

Principal Component Pursuit with Reduced Linear Measurements

Arvind Ganesh*, Kerui Min*, John Wright[†], and Yi Ma*,[‡]

* Dept. of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign

[†] Dept. of Electrical Engineering, Columbia University

[‡] Visual Computing Group, Microsoft Research Asia

Abstract

In this paper, we study the problem of decomposing a superposition of a low-rank matrix and a sparse matrix when a relatively few linear measurements are available. This problem arises in many data processing tasks such as aligning multiple images or rectifying regular texture, where the goal is to recover a low-rank matrix with a large fraction of corrupted entries in the presence of nonlinear domain transformation. We consider a natural convex heuristic to this problem which is a variant to the recently proposed Principal Component Pursuit. We prove that under suitable conditions, this convex program guarantees to recover the correct low-rank and sparse components despite reduced measurements. Our analysis covers both random and deterministic measurement models.

1 Introduction

Low-rank matrix recovery and approximation has been a popular area of research in many different fields. The popularity of low-rank matrices can be attributed to the fact that they arise in one of the most commonly used data models in real applications, namely when very high-dimensional data samples are assumed to lie approximately on a low-dimensional linear subspace. This model has been successfully employed in various problems such as face recognition [1], system identification [2], and information retrieval [3], for instance.

The most popular tool for low-rank matrix approximation is the Principal Component Analysis (PCA) [4, 5]. The basic idea of PCA is to find the “best low-rank approximation” (in an ℓ_2 -sense) to a given input matrix. Essentially, PCA finds a rank- r approximation to a given data matrix $D \in \mathbb{R}^{m \times n}$ by solving the following problem:

$$\min_L \|D - L\| \quad \text{s.t.} \quad \text{rank}(L) \leq r,$$

where $\|\cdot\|$ denotes the matrix spectral norm. It is well-known that the solution to this problem can be easily obtained by computing the Singular Value Decomposition (SVD) of D and retaining only the r largest singular values and the corresponding singular vectors. Besides the ease of computation, the PCA estimate has been shown to be optimal in the presence of isotropic Gaussian noise. However, the biggest drawback of PCA is that it breaks down even when one entry of the matrix is corrupted by an error of very large magnitude. Unfortunately, such large-magnitude, *non-Gaussian* errors often exist in real data. For instance, occlusions in images corrupt only a fraction of the pixels in an image, but the magnitude of corruption can be quite large.

There have been many works in the literature that try to make PCA robust to such gross, non-Gaussian errors and many models and solutions have been proposed. We here consider the specific problem of recovering a low-rank matrix $L_0 \in \mathbb{R}^{m \times n}$ from corrupted observations $D = L_0 + S_0$, where $S_0 \in \mathbb{R}^{m \times n}$ is a sparse matrix whose non-zero entries may have arbitrary magnitude. This problem has been studied in detail recently by various works in the literature [6, 7, 8]. It has been shown that under

rather broad conditions, the following convex program succeeds in recovering L_0 from D :

$$\min_{L, S} \|L\|_* + \lambda \|S\|_1 \quad \text{s.t.} \quad D = L + S, \quad (1)$$

where $\|\cdot\|_*$ denotes the nuclear norm¹, $\|\cdot\|_1$ denotes the ℓ_1 -norm², and $\lambda > 0$ is a weighting factor. This method has been dubbed Principal Component Pursuit (PCP) in [6]. In addition to being computationally tractable, it comes with very strong theoretical guarantees of recovery. Furthermore, follow-up works have shown that PCP is stable in the presence of additive Gaussian noise [9] and can recover L_0 even when the corruption matrix S_0 is not so sparse [10].

Besides being of theoretical interest, this convex optimization framework for low-rank matrix recovery has been employed very successfully to solve real problems in computer vision such as photometric stereo [11]. However, in practice, much more data, especially imagery data, can be viewed as low-rank only after some transformation is applied. For instance, an image of a building facade will become a low-rank matrix after the perspective distortion is rectified [12] or a set of face images of the same person will become linearly correlated only after they are properly aligned [13]. With our terminology here, we can write as $D \circ \tau = L_0 + S_0$ where τ belongs to certain transformation group. As the transformation τ is also unknown, one natural way to recover L_0 , S_0 and τ together is to approximate the nonlinear equation with its linearization at the current estimate of $\hat{\tau}$:

$$D \circ \hat{\tau} + \sum_{i=1}^p J_i d\tau_i = L + S,$$

where $\{J_i\}$ is the Jacobian of $D \circ \tau$ with respect to the parameters $\{\tau_i\}$ of τ . Then one can incrementally update the estimate for τ with $\hat{\tau} + d\tau$ by solving the following convex program:

$$\min_{L, S, d\tau_i} \|L\|_* + \lambda \|S\|_1 \quad \text{s.t.} \quad D + \sum_{i=1}^p J_i d\tau_i = L + S. \quad (2)$$

Empirically this scheme has been shown to work rather effectively in practice in both the image rectification problem [12] and the image alignment problem [13].

Although the convex program was proposed in the same spirit as PCP, we note that the linear constraint is different, and hence, the theoretical guarantees for PCP shown in [6, 7, 8] do not directly apply to this case. In this work, we attempt to fill the gap between theory and practice and try to understand under what conditions, the above extended version of PCP is expected to work correctly.

Let Q be the linear subspace in $\mathbb{R}^{m \times n}$ that is the orthogonal complement to the span of all the J_i 's, then its dimension is $q = mn - p$. Clearly, we can rewrite the above program in the following form:

$$\min_{L, S} \|L\|_* + \lambda \|S\|_1 \quad \text{s.t.} \quad \mathcal{P}_Q D = \mathcal{P}_Q (L + S), \quad (3)$$

where \mathcal{P}_Q is the orthogonal projection onto the linear subspace Q . Clearly, this program is a variation to PCP (1) in which the number of linear constraints has been reduced from mn to $q = mn - p$. Indeed, if Q is the entire space, then it reduces to the PCP. If Q is a linear subspace of matrices with support in $\Omega \subseteq [m] \times [n]$, then we have the special case of recovering L_0 from D , when only a subset of the entries in D are available. This case is akin to the low-rank matrix completion problem [14, 15, 16], and theoretical guarantees have been derived in [6, 17]. However, to the best of our knowledge, the case with a general subspace Q has not yet been analyzed in detail in the literature.

Our motivation to study when the convex program (3) succeeds with such reduced linear constraints is at least twofold. First, the relationships between Q and L_0 and S_0 will provide us better understanding about what type of images and signals for which techniques such as those used in [12, 13] are expected to work well. Second, we want to know how many general linear measurements we could reduce without

¹The sum of all singular values.

²The sum of absolute values of all matrix entries.

sacrificing the robustness of PCP for recovering the low-rank matrix L_0 . In these applications, the number of constraints reduced corresponds to the dimension of the transformation group. In the image rectification problem, the dimension of the transformation group p is typically fixed with respect to the size of the matrix; in the image alignment problem, however, the dimension typically grows linearly in m (or n). In either case, we need to know if the program (3) can tolerate up to a constant fraction of gross errors.

1.1 Notation

We first establish a set of notations that will be used throughout this work. We will assume that the matrices L_0 , S_0 and D in (3) have size $m \times n$. Without any loss of generality, we assume that $n \leq m$. We denote the rank of L_0 by r . Let $L_0 = U\Sigma V^*$ be the reduced Singular Value Decomposition (SVD) of L_0 . We define a linear subspace T as follows:

$$T \doteq \{UX^* + YV^* : X \in \mathbb{R}^{n \times r}, Y \in \mathbb{R}^{m \times r}\}. \quad (4)$$

Basically, T contains all matrices that share a common row space or column space with L_0 . We denote by Ω the support of S_0 . By a slight abuse of notation, we also represent by Ω the subspace of matrices whose support is contained in the support of S_0 . For any subspace $S \subseteq \mathbb{R}^{m \times n}$, $\mathcal{P}_S : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{m \times n}$ denotes the orthogonal projection operator onto S .

For any $X, Y \in \mathbb{R}^{m \times n}$, we define their inner product as $\langle X, Y \rangle = \text{trace}(X^*Y) = \sum_{ij} X_{ij}Y_{ij}$. We let $\|\cdot\|_F$ and $\|\cdot\|$ denote the matrix Frobenius norm and spectral norm, respectively. We also denote the ℓ_∞ -norm of a matrix X as $\|X\|_\infty = \max_{ij} |X_{ij}|$. We say that an event E occurs with high probability if $\mathbb{P}[E^c] \leq C m^{-\alpha}$, for some positive numerical constants C and α . Here, E^c denotes the event complement to E .

1.2 Main Assumptions

Obviously, successful recovery is not always guaranteed except under proper assumptions on the low-rank L_0 , sparse S_0 , and the subspace Q involved. For instance, if the matrix L_0 is itself a sparse matrix, then there is a fundamental ambiguity in the solution to be recovered. Here, we outline some of our assumptions that we will use throughout this paper. The assumptions we make here on L_0 and S_0 are essentially the same as those for PCP [6]. For completeness, we list them below.

We assume that each entry of the matrix belongs to the support of the sparse matrix S_0 independently with probability ρ . We denote this as $\text{supp}(S_0) \sim \text{Ber}(\rho)$. For simplicity, we assume the signs of the nonzero entries are also random.³ For the low-rank matrix L_0 , we assume the subspace T defined in (4) is incoherent to the standard basis (and hence the sparse matrix S_0). To be precise, let us denote the standard basis in \mathbb{R}^m and \mathbb{R}^n by \bar{e}_i and e_j , respectively, where $i \in [m]$ and $j \in [n]$. We assume (as in [14]) that

$$\max_{i \in [m]} \|U^* \bar{e}_i\|_2^2 \leq \frac{\mu r}{m}, \quad \max_{j \in [n]} \|V^* e_j\|_2^2 \leq \frac{\mu r}{n}, \quad \|UV^*\|_\infty \leq \sqrt{\frac{\mu r}{mn}}, \quad (5)$$

for some $\mu > 0$ and for all $(i, j) \in [m] \times [n]$. We recall that $r = \text{rank}(L_0)$. It follows from the above assumptions that for any $(i, j) \in [m] \times [n]$

$$\|\mathcal{P}_T \bar{e}_i e_j^*\|_F \leq \sqrt{\frac{2\mu r}{n}}. \quad (6)$$

Furthermore, it can be shown that $\|\mathcal{P}_{T^\perp} X\| \leq \|X\|$ for any $X \in \mathbb{R}^{m \times n}$.

In addition to the above assumptions, we define the following two properties of linear subspaces. We say that a linear subspace $S \subseteq \mathbb{R}^{m \times n}$ is

³The random sign assumption is not entirely necessary for obtaining the same qualitative results. One can follow the derandomization process in [6] to remove this assumption if needed.

- ν -coherent if there exists an orthonormal basis $\{G_i\}$ for S satisfying

$$\max_i \|G_i\|^2 \leq \frac{\nu}{n}. \quad (7)$$

- γ -constrained if

$$\max_{i,j} \|\mathcal{P}_S \bar{e}_i e_j^*\|_F^2 \leq \gamma. \quad (8)$$

If S is a random subspace, we say it is ν -coherent and γ -constrained if Eqns. (7) and (8) hold with high probability, respectively.

In this paper, we will deal with two different assumptions on the subspace Q as outlined below. We will see later that it is in fact convenient to make our assumptions on the subspace Q^\perp , rather than on Q itself. This is partly motivated from the model in (2) that was used in [12, 13], where the J_i 's are essentially a basis for Q^\perp . So, any assumptions on Q^\perp can be easily interpreted in terms of the J_i 's and this would help us make the connection to these applications more directly. We denote by p the dimension of the subspace Q^\perp .

- **Random subspace model.** Let $G_1, G_2, \dots, G_p \in \mathbb{R}^{m \times n}$ be an orthonormal basis for Q^\perp . We assume that this basis set is chosen uniformly at random from all possible orthonormal basis sets of size p in $\mathbb{R}^{m \times n}$. It can be shown that each of the G_i 's are identical in distribution to $H/\|H\|_F$, where the entries of $H \in \mathbb{R}^{m \times n}$ are i.i.d. according to a Gaussian distribution with mean 0 and variance $1/mn$.
- **Deterministic subspace model.** Under this model, we assume that Q^\perp is a fixed subspace which is ν -coherent, for some $\nu \geq 1$.

1.3 Main Results

With the above notation, we now briefly describe the main results we prove in this work. Although our results and proof methodology resemble those in [6], there are some important differences here. Particularly, we will see that the assumptions we make on the subspace Q greatly influences the kind of guarantees for recovery that can be derived.

As mentioned earlier, we will consider two different assumptions on the subspace Q . In the first one, we assume a *random subspace model* for Q^\perp . The main result that we prove in this work under this random subspace model is summarized as the following theorem.

Theorem 1 (Random Reduction). Fix any $C_p > 0$, and let Q^\perp be a p -dimensional random subspace of $\mathbb{R}^{m \times n}$ ($n \leq m$), L_0 a rank- r , μ -incoherent matrix, and $\text{supp}(S_0) \sim \text{Ber}(\rho)$. Then, provided that

$$r < C_r \frac{n}{\mu \log^2 m}, \quad p < C_p n, \quad \rho < \rho_0, \quad (9)$$

with high probability (L_0, S_0) is the unique optimal solution to (3) with $\lambda = m^{-1/2}$. Here, $C_r > 0$ and $\rho_0 \in (0, 1)$ are numerical constants.

Remark 1. In Theorem 1, “with high probability” means with probability at least $1 - \beta(C_p)m^{-c}$, with $c > 0$ numerical.

The scaling in this result covers several applications of interest: in [12], p is a fixed constant, while in [13], p scales linearly with n . Therefore, the above result already covers both these applications in terms of the number of reduced constraints. It states that with such reduced constraints, the convex program (3) can recover the low-rank matrix L_0 essentially under the same conditions as PCP. In particular, it can tolerate up to a constant fraction of errors.

In a work that is closely related to this one [18], we have shown that one can expect the convex program (3) to work under much more highly compressive scenario. More precisely, the dimension of

the subspace Q only needs to be on the order of $(mr + k) \log^2 m$ which is only a polylogarithmic factor more than the intrinsic degrees of freedom of the unknown L_0 and S_0 . One nice feature about the work of [18] is that the proof framework is very modular and the techniques are even applicable to more general structured signals beyond low-rank and sparse ones. Nevertheless, that result does not subsume the result here because in such highly compressive scenario, we cannot expect to tolerate error up to a constant fraction of the matrix entries. Obtaining the results in Theorem 1 and Theorem 2 seems to require arguments that are specially tailored to the PCP problem.

There is a common limitation for all results that are based on a random assumption for Q or Q^\perp : the random assumption does not hold in many real applications. For instance, in [12, 13], the subspace Q^\perp is typically spanned by a set of image Jacobians, which may not behave like random matrices. Therefore, it is desirable to have deterministic conditions on Q^\perp (or Q) that can be verified for the given data. We need theoretical guarantees for recovery when Q^\perp is a deterministic subspace. This is the second scenario that we will consider in this work, for which we have the following result:

Theorem 2 (Deterministic Reduction). Fix any $p \in \mathbb{Z}_+$, $\alpha \geq 1$, and $\nu \geq 1$. Then there exists $C_r > 0$ such that if Q^\perp is a ν -coherent p -dimensional subspace of $\mathbb{R}^{m \times n}$ ($n \leq m \leq \alpha n$), L_0 is a rank- r , μ -incoherent matrix, and $\text{supp}(S_0) \sim \text{Ber}(\rho)$, with high probability (L_0, S_0) is the unique optimal solution to (3) with $\lambda = m^{-1/2}$, provided that

$$r < C_r \min \left\{ \left(\frac{n}{\nu^2 p^2 \alpha} \right)^{1/2}, \left(\frac{n}{\alpha \nu \mu p} \right)^{1/3}, \frac{n}{\mu \log m} \right\}, \quad \rho < \rho_0, \quad (10)$$

where $C_r, \rho_0 \in (0, 1)$ are numerical constants.

Remark 2. Here, “with high probability” means with probability at least $1 - \beta(p, \alpha, \nu)m^{-c}$, with $c > 0$ numerical.

The ν -coherence condition essentially requires there exists an orthonormal basis for Q^\perp whose spectral norms are bounded above by $O(n^{-1/2})$. This is a condition that can be verified directly once the subspace Q or Q^\perp is given (say as the span of the Jacobians). This condition is also significantly weaker than the random subspace assumption in Theorem 1.

Because the assumptions are weaker, the orders of growth in Theorem 2, quite a bit more restrictive than those in Theorem 1. Nevertheless, this result can be very useful for the practical problems that we encountered in image rectification where the dimension of the transformation group is typically fixed (i.e. does not change with the matrix dimension). Theorem 2 suggests we should expect the program to work at least for deformation groups whose dimension is fixed. Although empirical results suggest that it could even grow as $O(n)$, we leave that for future investigation.

The remainder of this paper is organized as follows: In Section 2, we derive the optimality conditions for (L_0, S_0) to be the optimal solution to the convex program (3). In particular, we derive the conditions that a certain dual certificate must satisfy that would establish our main result. In Section 3, we provide a constructive procedure for the aforementioned dual certificate. In Section 4, we describe our main assumptions and the detailed steps of the proof of Theorem 1. In Section 5, we outline the proof of Theorem 2. Although the proof for both the deterministic case will follow a common strategy as the random case, there are a few important differences. In particular, we will highlight the parts where the proof deviates significantly from that of Theorem 1.

2 Existence of Dual Certificate

In this section, we prove the following lemma that establishes necessary and sufficient conditions for (L_0, S_0) to be the optimal solution to (3).

Lemma 1. Assume that $\dim(Q^\perp \oplus T \oplus \Omega) = \dim(Q^\perp) + \dim(T) + \dim(\Omega)$. (L_0, S_0) is the unique optimal solution to (3) if there exists a pair $(W, F) \in \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n}$ satisfying

$$UV^* + W = \lambda(\text{sgn}(S_0) + F) \in Q, \quad (11)$$

with $\mathcal{P}_T W = \mathbf{0}$, $\|W\| < 1$, $\mathcal{P}_\Omega F = \mathbf{0}$, and $\|F\|_\infty < 1$.

Proof. Consider a feasible solution to (3) of the form $(L_0 + H_L, S_0 - H_S)$. Clearly, we have that $\mathcal{P}_Q H_L = \mathcal{P}_Q H_S$. Under the conditions mentioned in the lemma, we will show that this pair does not minimize the cost function in (3), unless $H_L = H_S = \mathbf{0}$.

We first use the fact that $\|\cdot\|_*$ and $\|\cdot\|_1$ are convex functions. Consider any pair $(W_0, F_0) \in \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n}$ satisfying $\mathcal{P}_T W_0 = \mathbf{0}$, $\|W_0\| \leq 1$, $\mathcal{P}_\Omega F_0 = \mathbf{0}$, and $\|F_0\|_\infty \leq 1$. Then, $UV^* + W_0$ is a subgradient to $\|\cdot\|_*$ at L_0 , and $\text{sgn}(S_0) + F_0$ is a subgradient to $\|\cdot\|_1$ at S_0 . Therefore,

$$\|L_0 + H_L\|_* + \lambda\|S_0 - H_S\|_1 \geq \|L_0\|_* + \lambda\|S_0\|_1 + \langle UV^* + W_0, H_L \rangle - \lambda\langle \text{sgn}(S_0) + F_0, H_S \rangle.$$

By Hölder's inequality (and the duality of norms), it is possible to choose W_0 and F_0 such that

$$\langle W_0, H_L \rangle = \|\mathcal{P}_{T^\perp} H_L\|_*, \quad \langle F_0, H_S \rangle = -\|\mathcal{P}_{\Omega^\perp} H_S\|_1.$$

Then, we have

$$\begin{aligned} \|L_0 + H_L\|_* + \lambda\|S_0 - H_S\|_1 &\geq \|L_0\|_* + \lambda\|S_0\|_1 + \langle UV^*, H_L \rangle - \lambda\langle \text{sgn}(S_0), H_S \rangle \\ &\quad + \|\mathcal{P}_{T^\perp} H_L\|_* + \lambda\|\mathcal{P}_{\Omega^\perp} H_S\|_1. \end{aligned}$$

By assumption, we have

$$UV^* = \lambda(\text{sgn}(S_0) + F) - W,$$

with $\lambda(\text{sgn}(S_0) + F) \in Q$. Substituting for UV^* and using $\mathcal{P}_Q H_L = \mathcal{P}_Q H_S$, we get

$$\langle UV^*, H_L \rangle = \lambda\langle \text{sgn}(S_0), H_S \rangle + \lambda\langle F, H_S \rangle - \langle W, H_L \rangle.$$

Substituting this in the above inequality, we get

$$\begin{aligned} \|L_0 + H_L\|_* + \lambda\|S_0 - H_S\|_1 &\geq \|L_0\|_* + \lambda\|S_0\|_1 + \|\mathcal{P}_{T^\perp} H_L\|_* + \lambda\|\mathcal{P}_{\Omega^\perp} H_S\|_1 \\ &\quad + \lambda\langle F, H_S \rangle - \langle W, H_L \rangle. \end{aligned}$$

Let $\beta = \max\{\|W\|, \|F\|_\infty\} < 1$. Using Hölder's inequality, we get

$$\begin{aligned} \|L_0 + H_L\|_* + \lambda\|S_0 - H_S\|_1 &\geq \|L_0\|_* + \lambda\|S_0\|_1 + (1 - \beta)\|\mathcal{P}_{T^\perp}(H_L)\|_* \\ &\quad + (1 - \beta)\lambda\|\mathcal{P}_{\Omega^\perp}(H_S)\|_1. \end{aligned}$$

For non-zero H_L, H_S , the last term on the right hand side above can be zero only if $H_L \in T \setminus \{\mathbf{0}\}$ and $H_S \in \Omega \setminus \{\mathbf{0}\}$. Since $\Omega \cap T = \{\mathbf{0}\}$, $H_L \neq H_S$. We also have $\mathcal{P}_Q(H_L - H_S) = \mathbf{0}$. This implies that $H_L - H_S \in Q^\perp$, which is a contradiction since $Q^\perp \cap (T \oplus \Omega) = \{\mathbf{0}\}$. Thus, we have

$$\|L_0 + H_L\|_* + \lambda\|S_0 - H_S\|_1 > \|L_0\|_* + \lambda\|S_0\|_1,$$

for any non-zero *feasible* perturbation (H_L, H_S) . □

It is often convenient to relax the equality constraints on the dual certificate given in (11). Thus, similar to the proof outline in [6, 16], we now provide a slightly relaxed dual certificate condition.

Fact 1. Let S_1 and S_2 be two linear subspaces in $\mathbb{R}^{m \times n}$ with $S_1 \subseteq S_2$. Then, for any $X \in \mathbb{R}^{m \times n}$, we have $\mathcal{P}_{S_1} X = \mathcal{P}_{S_1} \mathcal{P}_{S_2} X$, and consequently, $\|\mathcal{P}_{S_1} X\|_F \leq \|\mathcal{P}_{S_2} X\|_F$.

Lemma 2. Suppose that $\dim(Q^\perp \oplus T \oplus \Omega) = \dim(Q^\perp) + \dim(T) + \dim(\Omega)$. Let $\Gamma = Q \cap T^\perp$ so that $\Gamma^\perp = Q^\perp \oplus T$. Assume that $\|\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp}\| < 1/2$ and $\lambda < 1$. Then, (L_0, S_0) is the unique optimal solution to (3) if there exists a pair $(W, F) \in \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n}$ satisfying

$$UV^* + W = \lambda(\text{sgn}(S_0) + F + \mathcal{P}_\Omega D) \in Q, \quad (12)$$

with $\mathcal{P}_T W = \mathbf{0}$, $\|W\| < 1/2$, $\mathcal{P}_\Omega F = \mathbf{0}$, $\|F\|_\infty < 1/2$, and $\|\mathcal{P}_\Omega D\|_F \leq 1/4$.

Proof. Proceeding along the same lines as in the proof of Lemma 1, for any feasible perturbation (H_L, H_S) , we get

$$\begin{aligned} \|L_0 + H_L\|_* + \lambda\|S_0 - H_S\|_1 &\geq \|L_0\|_* + \lambda\|S_0\|_1 + \frac{1}{2}\|\mathcal{P}_{T^\perp} H_L\|_* \\ &\quad + \frac{\lambda}{2}\|\mathcal{P}_{\Omega^\perp} H_S\|_1 + \lambda\langle \mathcal{P}_\Omega D, H_S \rangle \\ &\geq \|L_0\|_* + \lambda\|S_0\|_1 + \frac{1}{2}\|\mathcal{P}_{T^\perp} H_L\|_* \\ &\quad + \frac{\lambda}{2}\|\mathcal{P}_{\Omega^\perp} H_S\|_1 - \frac{\lambda}{4}\|\mathcal{P}_\Omega H_S\|_F. \end{aligned}$$

We note that

$$\begin{aligned} \|\mathcal{P}_\Omega H_S\|_F &\leq \|\mathcal{P}_\Omega \mathcal{P}_\Gamma H_S\|_F + \|\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp} H_S\|_F \\ &\leq \|\mathcal{P}_\Omega \mathcal{P}_\Gamma H_L\|_F + \frac{1}{2}\|H_S\|_F \\ &\leq \|\mathcal{P}_\Gamma H_L\|_F + \frac{1}{2}\|\mathcal{P}_\Omega H_S\|_F + \frac{1}{2}\|\mathcal{P}_{\Omega^\perp} H_S\|_F \\ &\leq \|\mathcal{P}_{T^\perp} H_L\|_F + \frac{1}{2}\|\mathcal{P}_\Omega H_S\|_F + \frac{1}{2}\|\mathcal{P}_{\Omega^\perp} H_S\|_F. \end{aligned}$$

In the second step above, we have used the fact that $\mathcal{P}_\Gamma H_L = \mathcal{P}_\Gamma H_S$ (since $\Gamma \subseteq Q$), and the final inequality follows from Fact 1. Thus, we have

$$\|\mathcal{P}_\Omega H_S\|_F \leq 2\|\mathcal{P}_{T^\perp} H_L\|_F + \|\mathcal{P}_{\Omega^\perp} H_S\|_F \leq 2\|\mathcal{P}_{T^\perp} H_L\|_* + \|\mathcal{P}_{\Omega^\perp} H_S\|_1.$$

Putting it all together, we get

$$\|L_0 + H_L\|_* + \lambda\|S_0 - H_S\|_1 \geq \|L_0\|_* + \lambda\|S_0\|_1 + \frac{1-\lambda}{2}\|\mathcal{P}_{T^\perp} H_L\|_* + \frac{\lambda}{4}\|\mathcal{P}_{\Omega^\perp} H_S\|_1.$$

The desired result follows from the fact that $Q^\perp \cap (T \oplus \Omega) = \{\mathbf{0}\}$. \square

3 Proof Strategy

By Lemma 2, in order for us to prove either Theorem 1 or 2, it is sufficient to produce a dual certificate $W \in \mathbb{R}^{m \times n}$ satisfying

$$\begin{cases} W \in T^\perp, \\ \mathcal{P}_{Q^\perp} W = -\mathcal{P}_{Q^\perp}(UV^*), \\ \|W\| < 1/2, \\ \|\mathcal{P}_\Omega(UV^* - \lambda \text{sgn}(S_0) + W)\|_F \leq \lambda/4, \\ \|\mathcal{P}_{\Omega^\perp}(UV^* + W)\|_\infty < \lambda/2. \end{cases} \quad (13)$$

To prove Theorems 1 and 2 under the above conditions, we try to construct the dual certificate W by following a similar strategy as that in the original PCP [6]. However, the extra projection of the observations onto the subspace Q adds significant difficulty to various technical parts of the proof. In

this section, we will outline the basic components for constructing such a certificate and then provide detailed proofs for each of the component in next sections. For simplicity, throughout our discussion below, we set $\Gamma \doteq Q \cap T^\perp$ so that $\Gamma^\perp = Q^\perp \oplus T$.

As the support of the sparse matrix is distributed as $\Omega \sim \text{Ber}(\rho)$ for some small $\rho \in (0, 1)$. This is, of course, equivalent to assuming that $\Omega^c \sim \text{Ber}(1 - \rho)$. Suppose that $\Omega_1, \Omega_2, \dots, \Omega_{j_0}$ are independent support sets such that $\Omega_j \sim \text{Ber}(q)$ for all j . Then, Ω^c and $\bigcup_{j=1}^{j_0} \Omega_j$ have the same probability distribution if $\rho = (1 - q)^{j_0}$. We now propose a construction for the dual certificate $W \doteq W^L + W^S + W^Q$ as follows. We use a combination of the golfing scheme proposed in [16] and the least norm approach.

1. *Construction of W^L using the golfing scheme.* Starting with $Y_0 = \mathbf{0}$, we iteratively define

$$Y_j = Y_{j-1} + q^{-1} \mathcal{P}_{\Omega_j} \mathcal{P}_{\Gamma^\perp} (UV^* - Y_{j-1}), \quad (14)$$

and set

$$W^L = \mathcal{P}_\Gamma Y_{j_0}, \quad (15)$$

where $j_0 = \lceil 2 \log m \rceil$.

2. *Construction of W^S by least norm solution.* We define W^S by the following least norm problem:

$$\begin{aligned} W^S &= \arg \min_X \|X\|_F \\ \text{subj. to} & \quad \mathcal{P}_\Omega X = \lambda \text{sgn}(S_0) \\ & \quad \mathcal{P}_{\Gamma^\perp} X = \mathbf{0}. \end{aligned} \quad (16)$$

3. *Construction of W^Q by least squares.* We define W^Q by the following least squares problem:

$$\begin{aligned} W^Q &= \arg \min_X \|X\|_F \\ \text{subj. to} & \quad \mathcal{P}_{Q^\perp} X = -\mathcal{P}_{Q^\perp} (UV^*) \\ & \quad \mathcal{P}_\Pi X = \mathbf{0}, \end{aligned} \quad (17)$$

where $\Pi = \Omega \oplus T$.

We note that under our assumptions (see Section 1.2), both the least squares programs above are feasible with high probability under both the random subspace model and the deterministic subspace model. This is because we will later show that the spectral norms of the linear operators $\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp}$ and $\mathcal{P}_{Q^\perp} \mathcal{P}_\Pi$ can be bounded below unity with high probability.

Thus, to prove that $W^L + W^S + W^Q$ is a valid dual certificate, we have to establish the following:

$$\|W^L + W^S + W^Q\| < 1/2, \quad (18)$$

$$\|\mathcal{P}_\Omega (UV^* + W^L)\|_F \leq \lambda/4, \quad (19)$$

$$\|\mathcal{P}_{\Omega^\perp} (UV^* + W^L + W^S + W^Q)\|_\infty < \lambda/2. \quad (20)$$

Lemma 3. Assume that $\Omega \sim \text{Ber}(\rho)$ for some small $\rho \in (0, 1)$ and the assumptions (5) and (7) hold true. Then, the matrix W^L obeys, with high probability,

1. $\|W^L\| < 1/4$,
2. $\|\mathcal{P}_\Omega (UV^* + W^L)\|_F < \lambda/4$,
3. $\|\mathcal{P}_{\Omega^\perp} (UV^* + W^L)\|_\infty < \lambda/4$.

Lemma 4. In addition to the assumptions in the previous lemma, assume that the signs of the non-zero entries of S_0 are i.i.d. random. Then, the matrix W^S obeys, with high probability,

1. $\|W^S\| < 1/8$,

$$2. \|\mathcal{P}_{\Omega^\perp} W^S\|_\infty < \lambda/8.$$

Lemma 5. Assume that $\Omega \sim \text{Ber}(\rho)$ for some small $\rho \in (0, 1)$ and the assumptions (5) and (7) hold true. Then, the matrix W^Q obeys, with high probability,

1. $\|W^Q\| < 1/8$,
2. $\|\mathcal{P}_{\Omega^\perp} W^Q\|_\infty < \lambda/8$.

The above lemmas together establish a valid dual certificate that satisfies Eqn. (18) to Eqn. (20).

4 Random Reduction: Proof of Theorem 1

In this section, we provide a detailed proof of Lemmas 3, 4, and 5 for the case when Q is a random subspace. Before proceeding to the main steps of the proof, we first establish some important properties and relationships among the different quantities involved in the problem.

4.1 Preliminaries

Lemma 6. Let Q^\perp be a linear subspace distributed according to the random subspace model described earlier. Then, for any $(i, j) \in [m] \times [n]$, with high probability,

$$\|\mathcal{P}_{Q^\perp} \bar{e}_i e_j^*\|_F \leq 4 \sqrt{\frac{p \log(mnp)}{mn}}. \quad (21)$$

Proof. For any $(i, j) \in [m] \times [n]$, we have

$$\|\mathcal{P}_{Q^\perp} \bar{e}_i e_j^*\|_F = \sqrt{\sum_{k=1}^p |\langle G_k, \bar{e}_i e_j^* \rangle|^2} \leq \sqrt{p} \max_k \|G_k\|_\infty. \quad (22)$$

We now derive a bound for $\|G_k\|_\infty$. Suppose that $M \in \mathbb{R}^{m \times n}$ is a random matrix whose entries are i.i.d. according to the standard normal distribution. Let us define

$$H = \frac{1}{\sqrt{mn}} M,$$

and $G = H/\|H\|_F$. Clearly, G is identical in distribution to G_1, G_2, \dots, G_p . We know that, for any $(i, j) \in [m] \times [n]$,

$$\mathbb{P}[|M_{ij}| > t] \leq \sqrt{\frac{2}{\pi}} \frac{e^{-t^2/2}}{t}.$$

Therefore, using a union bound, we get

$$\mathbb{P}[\|M\|_\infty > t] \leq \sqrt{\frac{2}{\pi}} \frac{mn}{t} e^{-t^2/2},$$

or equivalently,

$$\mathbb{P}\left[\|H\|_\infty > \frac{t}{\sqrt{mn}}\right] \leq \sqrt{\frac{2}{\pi}} \frac{mn}{t} e^{-t^2/2}.$$

Now, if we have p random matrices H_1, H_2, \dots, H_p , independent and identical in distribution to H , then

$$\mathbb{P}\left[\max_k \|H_k\|_\infty > \frac{t}{\sqrt{mn}}\right] \leq \sqrt{\frac{2}{\pi}} \frac{mnp}{t} e^{-t^2/2}.$$

Setting $t = \sqrt{4 \log(mnp)}$, we get

$$\mathbb{P} \left[\max_k \|H_k\|_\infty > \sqrt{\frac{4 \log(mnp)}{mn}} \right] \leq \sqrt{\frac{1}{2\pi}} \frac{1}{mnp \sqrt{\log(mnp)}}.$$

Thus, with high probability, we have that

$$\max_k \|H_k\|_\infty \leq \sqrt{\frac{4 \log(mnp)}{mn}}.$$

It can be shown that $\|H_k\|_F \geq 1/2$ with high probability. Thus, we have that

$$\max_k \|G_k\|_\infty \leq \sqrt{\frac{16 \log(mnp)}{mn}},$$

with high probability. The desired result follows from Eqn. (22). \square

Lemma 7. Assume that $p < mn/4$. Let Q^\perp be a linear subspace distributed according to the random subspace model. Then, with high probability, we have

$$\|\mathcal{P}_{Q^\perp} \mathcal{P}_T\| \leq 8 \left(\frac{\sqrt{p} + \sqrt{(m+n)r}}{\sqrt{mn}} \right). \quad (23)$$

Proof. Firstly, we note that Q^\perp is identical in distribution to a subspace spanned by p independent random matrices, each of whose entries are i.i.d. according to a Gaussian distribution with mean zero and variance $1/mn$. Let $\mathcal{H} : \mathbb{R}^p \rightarrow \mathbb{R}^{m \times n}$ be a linear operator defined as follows:

$$\mathcal{H}(\mathbf{x}) = \sum_{k=1}^p x_k H_k,$$

where the H_k 's are independent random matrices each of whose entries are i.i.d. according to a Gaussian distribution with mean zero and variance $1/mn$. Then, we have that \mathcal{P}_{Q^\perp} has the same distribution as the operator $\mathcal{H}(\mathcal{H}^* \mathcal{H})^{-1} \mathcal{H}^*$. Therefore, we have

$$\begin{aligned} & \mathbb{P} \left[\|\mathcal{P}_{Q^\perp} \mathcal{P}_T\| > 8 \left(\sqrt{\frac{p}{mn}} + \sqrt{\frac{(m+n)r}{mn}} \right) \right] \\ &= \mathbb{P} \left[\|\mathcal{H}(\mathcal{H}^* \mathcal{H})^{-1} \mathcal{H}^* \mathcal{P}_T\| > 8 \left(\sqrt{\frac{p}{mn}} + \sqrt{\frac{(m+n)r}{mn}} \right) \right] \\ &\leq \mathbb{P} \left[\|\mathcal{H}(\mathcal{H}^* \mathcal{H})^{-1}\| \|\mathcal{H}^* \mathcal{P}_T\| > 8 \left(\sqrt{\frac{p}{mn}} + \sqrt{\frac{(m+n)r}{mn}} \right) \right] \\ &\leq \mathbb{P} [\|\mathcal{H}(\mathcal{H}^* \mathcal{H})^{-1}\| > 4] + \mathbb{P} \left[\|\mathcal{H}^* \mathcal{P}_T\| > 2 \left(\sqrt{\frac{p}{mn}} + \sqrt{\frac{(m+n)r}{mn}} \right) \right]. \end{aligned}$$

Suppose that $R \in \mathbb{R}^{mn \times p}$ is a random matrix whose entries are i.i.d. according to a Gaussian distribution with mean zero and variance $1/mn$. It is easy to see that if we vectorize all the matrices, then R is the matrix analogue of the operator \mathcal{H} . Therefore, $\|\mathcal{H}(\mathcal{H}^* \mathcal{H})^{-1}\|$ has the same distribution as $(\sigma_{\min}(R))^{-1}$. Let $R' = \sqrt{mn} R$. Clearly, the entries of R' are i.i.d according to the standard normal distribution. Using the concentration results for 1-Lipschitz functions (see Proposition 2.18 in [19]) and the distribution of singular values of random Gaussian matrices [20], it is possible to show that

$$\mathbb{P} [\sigma_{\min}(R') \leq \sqrt{mn} - \sqrt{p} - t] \leq e^{-t^2/2},$$

for any $t \geq 0$. Consequently, we have that

$$\mathbb{P} \left[\sigma_{\min}(R) \leq 1 - \sqrt{\frac{p}{mn}} - t \right] \leq e^{-mnt^2/2}.$$

Setting $t = 1/4$ and by our assumption that $p < mn/4$, we get

$$\mathbb{P} \left[\sigma_{\min}(R) \leq \frac{1}{4} \right] = \mathbb{P} \left[\|\mathcal{H}(\mathcal{H}^*\mathcal{H})^{-1}\| \geq 4 \right] \leq e^{-mn/32}.$$

We now note that $\|\mathcal{H}^*\mathcal{P}_T\| = \|\mathcal{P}_T\mathcal{H}\|$ is identical in distribution to $\|M\|$, where $M \in \mathbb{R}^{(m+n)r \times p}$ is a random matrix whose entries are i.i.d. $\mathcal{N}(0, 1/mn)$. This is because the isotropic Gaussian distribution is rotation-invariant. Hence, without any loss of generality we can assume that the operator \mathcal{P}_T preserves only the first $\dim(T) = (m+n)r$ components of the basis elements H_1, \dots, H_p . Once again, invoking Proposition 2.18 in [19], we can show that

$$\mathbb{P} \left[\|M\| \geq \frac{\sqrt{p} + \sqrt{(m+n)r}}{\sqrt{mn}} + t \right] \leq e^{-mnt^2/2}.$$

Setting $t = \max \left\{ \sqrt{p/mn}, \sqrt{(m+n)r/mn} \right\}$, it follows that

$$\begin{aligned} & \mathbb{P} \left[\|M\| \geq 2 \left(\frac{\sqrt{p} + \sqrt{(m+n)r}}{\sqrt{mn}} \right) \right] \\ &= \mathbb{P} \left[\|\mathcal{H}^*\mathcal{P}_T\| \geq 2 \left(\frac{\sqrt{p} + \sqrt{(m+n)r}}{\sqrt{mn}} \right) \right] \\ &\leq \min \left\{ e^{-p/2}, e^{-(m+n)r/2} \right\}. \end{aligned}$$

Putting it all together, we get

$$\begin{aligned} & \mathbb{P} \left[\|\mathcal{P}_{Q^\perp}\mathcal{P}_T\| > 8 \left(\sqrt{\frac{p}{mn}} + \sqrt{\frac{(m+n)r}{mn}} \right) \right] \\ &\leq e^{-mn/32} + \min \left\{ e^{-p/2}, e^{-(m+n)r/2} \right\}. \end{aligned}$$

Thus, we have that

$$\|\mathcal{P}_{Q^\perp}\mathcal{P}_T\| \leq 8 \left(\frac{\sqrt{p} + \sqrt{(m+n)r}}{\sqrt{mn}} \right)$$

with high probability. \square

Lemma 8. Let Q^\perp be a linear subspace distributed according to the random subspace model and $\Omega \sim \text{Ber}(\rho)$. Then, with high probability, we have

$$\|\mathcal{P}_{Q^\perp}\mathcal{P}_\Omega\| \leq 8 \left(\sqrt{\frac{p}{mn}} + \sqrt{\frac{5\rho}{4}} \right). \quad (24)$$

Proof. Proceeding along the same lines of the proof of the previous lemma and conditioned on Ω , we get

$$\begin{aligned} & \mathbb{P} \left[\|\mathcal{P}_{Q^\perp}\mathcal{P}_\Omega\| > 8 \left(\sqrt{\frac{p}{mn}} + \sqrt{\frac{5\rho}{4}} \right) \mid |\Omega| \leq \frac{5}{4}\rho mn \right] \\ &\leq e^{-mn/32} + \min \left\{ e^{-p/2}, e^{-5mn\rho/8} \right\}. \end{aligned}$$

Using Bernstein's inequality, it is possible to show that

$$\mathbb{P}[|\Omega| > mn\rho(1+\delta)] \leq 2 \exp\left(-\frac{mn\rho\delta^2}{1-\rho+\frac{2\delta}{3}}\right) \leq 2 \exp\left(-\frac{3}{5}mn\rho\delta^2\right),$$

for any $\delta \in (0, 1)$. We set $\delta = 1/4$. Thus, we have

$$\begin{aligned} & \mathbb{P}\left[\|\mathcal{P}_{Q^\perp}\mathcal{P}_\Omega\| > 8\left(\sqrt{\frac{p}{mn}} + \sqrt{\frac{5\rho}{4}}\right)\right] \\ & \leq \mathbb{P}\left[\|\mathcal{P}_{Q^\perp}\mathcal{P}_\Omega\| > 8\left(\sqrt{\frac{p}{mn}} + \sqrt{\frac{5\rho}{4}}\right) \mid |\Omega| \leq \frac{5}{4}\rho mn\right] + \mathbb{P}\left[|\Omega| > \frac{5}{4}\rho mn\right] \\ & \leq e^{-mn/32} + \min\left\{e^{-p/2}, e^{-5mn\rho/8}\right\} + 2e^{-3mn\rho/80}. \end{aligned}$$

Thus, we have that

$$\|\mathcal{P}_{Q^\perp}\mathcal{P}_\Omega\| \leq 8\left(\sqrt{\frac{p}{mn}} + \sqrt{\frac{5\rho}{4}}\right)$$

with high probability. \square

Lemma 9. Let $\Omega \sim \text{Ber}(\rho)$. Then, with high probability,

$$\|\mathcal{P}_\Omega\mathcal{P}_T\|^2 \leq \rho + \epsilon, \quad (25)$$

provided that $1 - \rho \geq C_0 \epsilon^{-2} \frac{\mu r \log m}{n}$ for some numerical constant $C_0 > 0$.

Proof. See Corollary 2.7 in [6]. \square

We now prove the following two results that would help us establish incoherence relations with subspaces obtained by a direct sum of two incoherent subspaces.

Lemma 10. Let S_1 and S_2 be any two linear subspaces in $\mathbb{R}^{m \times n}$ satisfying $\|\mathcal{P}_{S_1}\mathcal{P}_{S_2}\| \leq \alpha < 1$. We define $S = S_1 \oplus S_2$. Then, for any $X \in \mathbb{R}^{m \times n}$, we have

$$\|\mathcal{P}_S X\|_F^2 \leq (1 - \alpha)^{-1} (\|\mathcal{P}_{S_1} X\|_F^2 + \|\mathcal{P}_{S_2} X\|_F^2). \quad (26)$$

Proof. We denote by $\text{vec} : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{mn}$, the operation of converting a matrix to a vector by stacking its columns one below another. Suppose that d_1 and d_2 are the dimensions of the subspaces S_1 and S_2 , respectively. Then, there exist matrices $B_1 \in \mathbb{R}^{mn \times d_1}$ and $B_2 \in \mathbb{R}^{mn \times d_2}$ whose columns constitute orthonormal bases for S_1 and S_2 , respectively.

Let $M \doteq [B_1 \ B_2]$. Clearly, the columns of M constitute a basis for the subspace S in \mathbb{R}^{mn} . Hence, for any $X \in \mathbb{R}^{m \times n}$, its projection onto S can be expressed as follows:

$$\text{vec}(\mathcal{P}_S X) = M(M^*M)^{-1}M^*\text{vec}(X).$$

We note that $\|B_1^*\text{vec}(X)\|_2 = \|\mathcal{P}_{S_1} X\|_F$ and $\|B_2^*\text{vec}(X)\|_2 = \|\mathcal{P}_{S_2} X\|_F$. Therefore, we have

$$\begin{aligned} \|\mathcal{P}_S X\|_F^2 &= \|\text{vec}(\mathcal{P}_S X)\|_2^2 \\ &= \|M(M^*M)^{-1}M^*\text{vec}(X)\|_2^2 \\ &\leq \|M(M^*M)^{-1}\|^2 \cdot \|M^*\text{vec}(X)\|_2^2 \\ &= \|M(M^*M)^{-1}\|^2 \cdot (\|\mathcal{P}_{S_1} X\|_F^2 + \|\mathcal{P}_{S_2} X\|_F^2) \end{aligned}$$

Let $M^\dagger \doteq (M^*M)^{-1}M^*$ denote the Moore-Penrose pseudoinverse of M . It is evident that $\|M^\dagger\| = \|M(M^*M)^{-1}\|$. But we know that $\|M^\dagger\| = (\sigma_{\min}(M))^{-1}$, where $\sigma_{\min}(M)$ is the smallest non-zero singular value of M . Using the fact that B_1 and B_2 have orthonormal columns, we can show that $(\sigma_{\min}(M))^2 = \lambda_{\min}(M^*M) \geq 1 - \alpha$, where $\lambda_{\min}(M^*M)$ is the smallest eigenvalue of M^*M .⁴ Therefore, we have

$$\begin{aligned} \|\mathcal{P}_S X\|_F^2 &\leq (\sigma_{\min}(M))^{-2} (\|\mathcal{P}_{S_1} X\|_F^2 + \|\mathcal{P}_{S_2} X\|_F^2) \\ &\leq (1 - \alpha)^{-1} (\|\mathcal{P}_{S_1} X\|_F^2 + \|\mathcal{P}_{S_2} X\|_F^2). \end{aligned}$$

□

Suppose that $\|\mathcal{P}_{Q^\perp} \mathcal{P}_T\| < 1/2$.⁵ Then, it follows that

$$\|\mathcal{P}_{\Gamma^\perp} \bar{e}_i e_j^*\|_F^2 \leq 4 \left(\frac{8p \log(mnp)}{mn} + \frac{\mu r}{n} \right), \quad (27)$$

with high probability, for all $(i, j) \in [m] \times [n]$. In other words, with high probability, when Q^\perp is distributed according to the random subspace model, we have that the subspace Γ^\perp is γ -constrained with $\gamma = 4 \left(\frac{8p \log(mnp)}{mn} + \frac{\mu r}{n} \right)$. We further note that $\gamma \log m = O(1/\log m)$ under the conditions of Theorem 1. This fact will be used frequently in our proof below.

Lemma 11. Let S_1, S_2 and S_3 be any three linear subspaces in $\mathbb{R}^{m \times n}$ satisfying $\dim(S_1 \oplus S_2 \oplus S_3) = \dim(S_1) + \dim(S_2) + \dim(S_3)$, and $\|\mathcal{P}_{S_1} \mathcal{P}_{S_2}\| \leq \alpha_{1,2} < 1$, $\|\mathcal{P}_{S_2} \mathcal{P}_{S_3}\| \leq \alpha_{2,3} < 1$ and $\|\mathcal{P}_{S_3} \mathcal{P}_{S_1}\| \leq \alpha_{3,1} < 1$. We define $S = S_1 \oplus S_2$. Then, we have

$$\|\mathcal{P}_S \mathcal{P}_{S_3}\| \leq \sqrt{\frac{\alpha_{2,3}^2 + \alpha_{3,1}^2}{1 - \alpha_{1,2}}}. \quad (28)$$

Proof. The proof is a simple application of Lemma 10. We note that, for any $X \in \mathbb{R}^{m \times n}$,

$$\begin{aligned} \|\mathcal{P}_S \mathcal{P}_{S_3} X\|_F^2 &\leq (1 - \alpha_{1,2})^{-1} (\|\mathcal{P}_{S_1} \mathcal{P}_{S_3} X\|_F^2 + \|\mathcal{P}_{S_2} \mathcal{P}_{S_3} X\|_F^2) \\ &\leq (1 - \alpha_{1,2})^{-1} (\|\mathcal{P}_{S_1} \mathcal{P}_{S_3}\|^2 + \|\mathcal{P}_{S_2} \mathcal{P}_{S_3}\|^2) \|X\|_F^2 \\ &\leq (1 - \alpha_{1,2})^{-1} (\alpha_{3,1}^2 + \alpha_{2,3}^2) \|X\|_F^2. \end{aligned}$$

It follows that

$$\|\mathcal{P}_S \mathcal{P}_{S_3}\| \leq \sqrt{\frac{\alpha_{2,3}^2 + \alpha_{3,1}^2}{1 - \alpha_{1,2}}}.$$

□

Lemma 12. Let $\Omega \sim \text{Ber}(\rho)$ and Γ^\perp be γ -constrained. Then, for any $\epsilon \in (0, 1)$, with high probability,

$$\|\mathcal{P}_{\Gamma^\perp} - \rho^{-1} \mathcal{P}_{\Gamma^\perp} \mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp}\| \leq \epsilon, \quad (29)$$

provided that $\rho \geq C \cdot \epsilon^{-2} \gamma \log m$ for some numerical constant $C > 0$.

Proof. The proof is very similar to that of Theorem 4.1 in [14]. We highlight the main steps here. For each $(i, j) \in [m] \times [n]$, we define binary random variables δ_{ij} , each takes value 1 if $(i, j) \in \Omega$, and 0 otherwise. We note that

$$\begin{aligned} \mathcal{P}_{\Gamma^\perp} \mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp} &= \sum_{ij} \delta_{ij} \mathcal{P}_{\Gamma^\perp} \bar{e}_i e_j^* \otimes \mathcal{P}_{\Gamma^\perp} \bar{e}_i e_j^*, \\ \mathbb{E}[\mathcal{P}_{\Gamma^\perp} \mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp}] &= \rho \mathcal{P}_{\Gamma^\perp}, \end{aligned}$$

⁴Since M has full column rank, M^*M is positive definite.

⁵From Lemma 7 and the assumptions of Theorem 1, this is true with high probability for sufficiently large m, n .

where \otimes denotes the outer or tensor product between matrices. Applying a concentration result for operators of the above form, as established in [21], we have, with high probability,

$$\|\mathcal{P}_{\Gamma^\perp} - \rho^{-1} \mathcal{P}_{\Gamma^\perp} \mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp}\| \leq C' \sqrt{\frac{\log(mn)}{\rho}} \max_{ij} \|\mathcal{P}_{\Gamma^\perp} \bar{e}_i e_j^*\|_F \quad (30)$$

$$\leq C' \sqrt{\frac{\gamma \log(mn)}{\rho}}, \quad (31)$$

provided that the right hand side is smaller than 1. Here, $C' > 0$ is a numerical constant. The desired result follows by noting that $n \leq m$, and bounding the right hand side by $\epsilon \in (0, 1)$. \square

Lemma 13. Let $Z \in \Gamma^\perp$ be fixed, Γ^\perp be γ -constrained, and $\Omega \sim \text{Ber}(\rho)$. Then, with high probability,

$$\|Z - \rho^{-1} \mathcal{P}_{\Gamma^\perp} \mathcal{P}_\Omega Z\|_\infty \leq \epsilon \|Z\|_\infty, \quad (32)$$

provided that $\rho \geq C_0 \cdot \epsilon^{-2} \gamma \log m$ for some numerical constant $C_0 > 64/3$.

Proof. Let δ_{ij} be a sequence of independent Bernoulli random variables such that

$$\delta_{ij} = \begin{cases} 1, & \text{if } (i, j) \in \Omega, \\ 0, & \text{otherwise.} \end{cases}$$

We define $Z' \doteq Z - \rho^{-1} \mathcal{P}_{\Gamma^\perp} \mathcal{P}_\Omega Z$. Then,

$$Z' = \sum_{ij} (1 - \rho^{-1} \delta_{ij}) Z_{ij} \mathcal{P}_{\Gamma^\perp} \bar{e}_i e_j^*.$$

For any $(i_0, j_0) \in [m] \times [n]$, we can express $Z'_{i_0 j_0}$ as a sum of independent random variables as shown below:

$$Z'_{i_0 j_0} = \sum_{ij} R_{ij}, \quad R_{ij} = (1 - \rho^{-1} \delta_{ij}) Z_{ij} \langle \mathcal{P}_{\Gamma^\perp} \bar{e}_i e_j^*, \bar{e}_{i_0} e_{j_0}^* \rangle.$$

It is easy to show that the R_{ij} 's are zero-mean random variables with variance given by

$$\text{Var}(R_{ij}) = (1 - \rho) \rho^{-1} |Z_{ij}|^2 |\langle \mathcal{P}_{\Gamma^\perp} \bar{e}_i e_j^*, \bar{e}_{i_0} e_{j_0}^* \rangle|^2.$$

Therefore,

$$\begin{aligned} \sum_{ij} \text{Var}(R_{ij}) &= (1 - \rho) \rho^{-1} \sum_{ij} |Z_{ij}|^2 |\langle \mathcal{P}_{\Gamma^\perp} \bar{e}_i e_j^*, \bar{e}_{i_0} e_{j_0}^* \rangle|^2 \\ &\leq (1 - \rho) \rho^{-1} \|Z\|_\infty^2 \sum_{ij} |\langle \bar{e}_i e_j^*, \mathcal{P}_{\Gamma^\perp} \bar{e}_{i_0} e_{j_0}^* \rangle|^2 \\ &= (1 - \rho) \rho^{-1} \|Z\|_\infty^2 \|\mathcal{P}_{\Gamma^\perp} \bar{e}_{i_0} e_{j_0}^*\|_F^2 \\ &\leq (1 - \rho) \rho^{-1} \gamma \|Z\|_\infty^2, \end{aligned}$$

where the last inequality holds with high probability. Furthermore, we have

$$\begin{aligned} |R_{ij}| &\leq \rho^{-1} \|Z\|_\infty |\langle \mathcal{P}_{\Gamma^\perp} \bar{e}_i e_j^*, \bar{e}_{i_0} e_{j_0}^* \rangle| \\ &\leq \rho^{-1} \|Z\|_\infty \|\mathcal{P}_{\Gamma^\perp} \bar{e}_i e_j^*\|_F \|\mathcal{P}_{\Gamma^\perp} \bar{e}_{i_0} e_{j_0}^*\|_F \\ &\leq \rho^{-1} \gamma \|Z\|_\infty, \end{aligned}$$

with high probability. Thus, using Bernstein's inequality, we obtain

$$\mathbb{P} [|Z'_{i_0 j_0}| > \epsilon \|Z\|_\infty] \leq 2 \exp \left(- \frac{\epsilon^2 \rho}{2\gamma (\frac{\epsilon}{3} + 1 - \rho)} \right).$$

Choosing $\epsilon < 1$, we can reduce the above expression to

$$\mathbb{P} [|Z'_{i_0 j_0}| > \epsilon \|Z\|_\infty] \leq 2 \exp \left(-\frac{3\epsilon^2 \rho}{8\gamma} \right).$$

If $\rho \geq C_0 \epsilon^{-2} \gamma \log m$ for some numerical constant $C_0 > 64/3$, then we have

$$\mathbb{P} [|Z'_{i_0 j_0}| > \epsilon \|Z\|_\infty] \leq 2 \exp \left(-\frac{3C_0 \log m}{32} \right).$$

Applying a union bound, we get

$$\begin{aligned} \mathbb{P} [\|Z'\|_\infty > \epsilon \|Z\|_\infty] &\leq 2mn \exp \left(-\frac{3C_0 \log m}{32} \right) \\ &\leq 2m^{(2-\frac{3C_0}{32})}. \end{aligned} \tag{33}$$

Since $C_0 > 64/3$, we obtain the desired result. \square

The following lemma is a restatement of Theorem 6.3 in [14].

Lemma 14. Let $Z \in \mathbb{R}^{m \times n}$ be fixed, and $\Omega \sim \text{Ber}(\rho)$. Then, with high probability,

$$\|Z - \rho^{-1} \mathcal{P}_\Omega Z\| \leq C'_0 \sqrt{\frac{m \log m}{\rho}} \|Z\|_\infty, \tag{34}$$

provided that $\rho \geq C'_0 \frac{\log m}{n}$, where $C'_0 > 0$ is a numerical constant.

4.2 Proof of Lemma 3

Before proceeding to the actual proof, we introduce some additional notation. Let $Z_j \doteq UV^* - \mathcal{P}_{\Gamma^\perp} Y_j$, where Y_j 's are defined in Eqn. (14). Evidently, $Z_j \in \Gamma^\perp$ for all $j \geq 0$. The recursive relation between the Y_j 's can then be expressed as

$$Z_j = (\mathcal{P}_{\Gamma^\perp} - q^{-1} \mathcal{P}_{\Gamma^\perp} \mathcal{P}_{\Omega_j} \mathcal{P}_{\Gamma^\perp}) Z_{j-1}, \quad Z_0 = UV^*. \tag{35}$$

Let us assume that $\epsilon \in (0, e^{-1})$. From Lemma 13, we have that

$$\|Z_j\|_\infty \leq \epsilon \|Z_{j-1}\|_\infty,$$

with high probability, provided that

$$q \geq C_0 \epsilon^{-2} \gamma \log m, \tag{36}$$

where $C_0 > 64/3$ is a numerical constant. Since $Z_0 = UV^*$, with high probability, we have

$$\begin{aligned} \|Z_j\|_\infty &\leq \epsilon^j \|UV^*\|_\infty \\ &\leq \epsilon^j \sqrt{\frac{\mu r}{mn}}. \end{aligned}$$

The second inequality above follows from our assumptions about the matrices U and V . Furthermore, when Eqn. (36) holds, we also have, with high probability,

$$\|Z_j\|_F \leq \epsilon \|Z_{j-1}\|_F \tag{37}$$

using Lemma 12. Once again, since $Z_0 = UV^*$, we deduce that

$$\begin{aligned} \|Z_j\|_F &\leq \epsilon^j \|UV^*\|_F \\ &= \epsilon^j \sqrt{r} \end{aligned} \tag{38}$$

with high probability.

4.2.1 Bounding $\|W^L\|$

We first introduce a few notions before deriving a bound on $\|W^L\|$. We let R denote the linear subspace obtained by projecting all the points in Q^\perp onto T^\perp . By a slight abuse of notation, we denote this by

$$R = \mathcal{P}_{T^\perp} Q^\perp. \quad (39)$$

We note that if Q^\perp is a random p -dimensional subspace in $\mathbb{R}^{m \times n}$, then with probability one, R is a p -dimensional subspace of T^\perp . It is easy to verify that for any $X \in \mathbb{R}^{m \times n}$, we have

$$\mathcal{P}_{\Gamma^\perp} X = \mathcal{P}_T X + \mathcal{P}_R X.$$

We note that

$$Y_{j_0} = \sum_{j=1}^{j_0} q^{-1} \mathcal{P}_{\Omega_j} Z_{j-1}. \quad (40)$$

Thus, we have

$$\begin{aligned} \|W^L\| &= \|\mathcal{P}_\Gamma Y_{j_0}\| \\ &\leq \sum_{j=1}^{j_0} \|q^{-1} \mathcal{P}_\Gamma \mathcal{P}_{\Omega_j} \mathcal{P}_{\Gamma^\perp} Z_{j-1}\| \\ &= \sum_{j=1}^{j_0} \|\mathcal{P}_\Gamma (q^{-1} \mathcal{P}_{\Omega_j} - \mathcal{I}) \mathcal{P}_{\Gamma^\perp} Z_{j-1}\| \\ &\leq \sum_{j=1}^{j_0} \|\mathcal{P}_{\Gamma^\perp} (q^{-1} \mathcal{P}_{\Omega_j} - \mathcal{I}) \mathcal{P}_{\Gamma^\perp} Z_{j-1}\| + \sum_{j=1}^{j_0} \|(q^{-1} \mathcal{P}_{\Omega_j} - \mathcal{I}) \mathcal{P}_{\Gamma^\perp} Z_{j-1}\|. \end{aligned}$$

The second term in the above inequality can be bounded with high probability using Lemma 14 as follows:

$$\begin{aligned} \sum_{j=1}^{j_0} \|(q^{-1} \mathcal{P}_{\Omega_j} - \mathcal{I}) \mathcal{P}_{\Gamma^\perp} Z_{j-1}\| &\leq C'_0 \sqrt{\frac{m \log m}{q}} \sum_{j=1}^{j_0} \|Z_{j-1}\|_\infty \\ &\leq C'_0 \sqrt{\frac{m \log m}{q}} \sum_{j=1}^{j_0} \epsilon^{j-1} \sqrt{\frac{\mu r}{mn}} \\ &\leq C'_0 \sqrt{\frac{\mu r \log m}{q n}} (1 - \epsilon)^{-1}, \end{aligned}$$

provided that

$$q \geq \max \left\{ C'_0 \frac{\log m}{n}, C_0 \epsilon^{-2} \gamma \log m \right\}.$$

On the other hand, each term in the summation in the first term can be split as

$$\begin{aligned} &\|\mathcal{P}_{\Gamma^\perp} (q^{-1} \mathcal{P}_{\Omega_j} - \mathcal{I}) \mathcal{P}_{\Gamma^\perp} Z_{j-1}\| \\ &\leq \|\mathcal{P}_T (q^{-1} \mathcal{P}_{\Omega_j} - \mathcal{I}) \mathcal{P}_{\Gamma^\perp} Z_{j-1}\| + \|\mathcal{P}_R (q^{-1} \mathcal{P}_{\Omega_j} - \mathcal{I}) \mathcal{P}_{\Gamma^\perp} Z_{j-1}\| \\ &\leq 2 \|(q^{-1} \mathcal{P}_{\Omega_j} - \mathcal{I}) \mathcal{P}_{\Gamma^\perp} Z_{j-1}\| + \|\mathcal{P}_R (q^{-1} \mathcal{P}_{\Omega_j} - \mathcal{I}) Z_{j-1}\|. \end{aligned}$$

We have already seen how the first term in the above inequality can be bounded with high probability. Hence, we now focus on the second term. We first state the matrix Bernstein inequality (see Theorem 1.4 in [22]) that will enable us to derive a bound on the second term.

Theorem 3 (Matrix Bernstein Inequality). Let $M_1, \dots, M_k \in \mathbb{R}^{d_1 \times d_2}$ be k independent random matrices satisfying

$$\mathbb{E}[M_i] = \mathbf{0}, \quad \|M_i\| \leq S \text{ almost surely, } i = 1, \dots, k. \quad (41)$$

We set

$$\sigma^2 = \max \left\{ \left\| \sum_{i=1}^k \mathbb{E}[M_i^* M_i] \right\|, \left\| \sum_{i=1}^k \mathbb{E}[M_i M_i^*] \right\| \right\}. \quad (42)$$

Then, for any $t > 0$, we have

$$\mathbb{P} \left[\left\| \sum_{i=1}^k M_i \right\| > t \right] \leq (d_1 + d_2) \exp \left(-\frac{t^2}{2\sigma^2 + 3St} \right). \quad (43)$$

Using Theorem 3, we will now show that, with high probability,

$$\|\mathcal{P}_R(q^{-1}\mathcal{P}_{\Omega_j} - \mathcal{I})Z_{j-1}\| \leq \tilde{C}p\sqrt{m} \log m \|Z_{j-1}\|_\infty.$$

The proof is as follows.

For every $(i, l) \in [m] \times [n]$, let us define $M_{il} \doteq H_{il}(Z_{j-1})_{il} \mathcal{P}_R \bar{e}_i e_l^*$, where the H_{il} 's are independent random variables distributed as follows:

$$H_{il} = \begin{cases} 1, & \text{w.p. } 1 - q \\ 1 - q^{-1}, & \text{w.p. } q \end{cases}$$

We note that $\sum_{i,l} M_{il}$ has the same distribution as $\mathcal{P}_R(\mathcal{I} - q^{-1}\mathcal{P}_{\Omega_j})Z_{j-1}$. Since the H_{il} 's are independent zero-mean random variables that are independent of Z_{j-1} , we have that, for any $(i, l) \in [m] \times [n]$,

$$\mathbb{E}[M_{il} | Z_{j-1}] = \mathbf{0}.$$

We record two useful bounds. We have that

$$1 - \rho = \mathbb{P}[\cup_j \{(i, l) \in \Omega_j\}] \leq j_0 q. \quad (44)$$

So $q \geq (1 - \rho)/j_0$. Since $|H_{il}| \leq q^{-1}$ almost surely, and $j_0 \geq C/\log m$, we have

$$|H_{il}| \leq O(\log m) \text{ almost surely.} \quad (45)$$

We also have

$$\|\mathcal{P}_R \bar{e}_i e_l^*\| \leq \|\mathcal{P}_R \bar{e}_i e_l^*\|_F \leq 1, \quad (46)$$

for any $(i, l) \in [m] \times [n]$. It follows that $\|M_{il}\| \leq O(\log m) \|Z_{j-1}\|_\infty$ almost surely.

Now we bound the variance term. It can be shown that $\mathbb{E}[H_{il}^2] = O(\log m)$. Let B_1, \dots, B_p be such an orthonormal basis for R . Then, we have

$$\left\| \sum_{il} \mathbb{E}[M_{il} M_{il}^*] \right\| = \left\| \sum_{il} \mathbb{E}[H_{il}^2] \mathcal{P}_R [\bar{e}_i e_l^*] (\mathcal{P}_R [\bar{e}_i e_l^*])^* (Z_{j-1})_{il}^2 \right\| \quad (47)$$

$$\leq O(\log m) \|Z_{j-1}\|_\infty^2 \left\| \sum_{il} \mathcal{P}_R [\bar{e}_i e_l^*] (\mathcal{P}_R [\bar{e}_i e_l^*])^* \right\| \quad (48)$$

$$= O(\log m) \|Z_{j-1}\|_\infty^2 \left\| \sum_{il} \left(\sum_{s=1}^p B_s \langle B_s, \bar{e}_i e_l^* \rangle \right) \left(\sum_{t=1}^p B_t \langle B_t, \bar{e}_i e_l^* \rangle \right)^* \right\| \quad (49)$$

$$= O(\log m) \|Z_{j-1}\|_\infty^2 \left\| \sum_t B_t B_t^* \right\| \quad (50)$$

$$\leq O(\log m) p \|Z_{j-1}\|_\infty^2 \quad (51)$$

A similar bound holds for the other variance term $\mathbb{E}[M_{il}^* M_{il}]$.

Now, using the matrix Bernstein inequality, we have

$$\begin{aligned} & \mathbb{P}(\|\mathcal{P}_R(q^{-1}\mathcal{P}_{\Omega_j} - \mathcal{I})Z_{j-1}\| > t | Z_{j-1}, Q) \\ & \leq (m+n) \exp\left(-\frac{t^2}{C_1 p \log m \|Z_{j-1}\|_\infty^2 + C_2 \log m \|Z_{j-1}\|_\infty t}\right). \end{aligned}$$

Therefore, removing the conditioning, we have that, with high probability,

$$\|\mathcal{P}_R(q^{-1}\mathcal{P}_{\Omega_j} - \mathcal{I})Z_{j-1}\| \leq \tilde{C}\sqrt{m} \log m \|Z_{j-1}\|_\infty,$$

for any j , for some numerical constant $\tilde{C} > 0$.

Thus we have that, with high probability,

$$\begin{aligned} \sum_{j=1}^{j_0} \|\mathcal{P}_R(q^{-1}\mathcal{P}_{\Omega_j} - \mathcal{I})Z_{j-1}\| & \leq \sum_{j=1}^{j_0} \tilde{C}\sqrt{m} \log m \|Z_{j-1}\|_\infty \\ & \leq \tilde{C}\sqrt{m} \log m \sqrt{\frac{\mu r}{mn}} (1-\epsilon)^{-1} \\ & \leq \tilde{C} \log m \sqrt{\frac{\mu r}{n}} (1-\epsilon)^{-1}. \end{aligned}$$

Under the assumptions of Theorem 1, the bound on the right hand side can be made arbitrarily small. This gives us the desired bound.

4.2.2 Bounding $\|\mathcal{P}_\Omega(UV^* + W^L)\|_F$

We now prove the second part of Lemma 3. First, we note that $\mathcal{P}_\Omega Y_{j_0} = \mathbf{0}$ by construction. Therefore,

$$\mathcal{P}_\Omega(UV^* + \mathcal{P}_\Gamma Y_{j_0}) = \mathcal{P}_\Omega(UV^* - \mathcal{P}_{\Gamma^\perp} Y_{j_0}) = \mathcal{P}_\Omega Z_{j_0}. \quad (52)$$

Consequently, we have

$$\begin{aligned} \|\mathcal{P}_\Omega(UV^* + \mathcal{P}_\Gamma Y_{j_0})\|_F & = \|\mathcal{P}_\Omega Z_{j_0}\|_F \\ & \leq \|Z_{j_0}\|_F \\ & \leq \epsilon^{j_0} \sqrt{r} \\ & \leq \frac{\sqrt{r}}{m^2}. \end{aligned}$$

The last step follows from the fact that $\epsilon < e^{-1}$ and $j_0 \geq 2 \log m$.

4.2.3 Bounding $\|\mathcal{P}_{\Omega^\perp}(UV^* + W^L)\|_\infty$

We now prove the final part of Lemma 3. We note that

$$UV^* + W^L = UV^* + \mathcal{P}_\Gamma Y_{j_0} = Y_{j_0} + Z_{j_0}. \quad (53)$$

Since we have already proved that $\|Z_{j_0}\|_F < \lambda/8$, it is sufficient to show that $\|Y_{j_0}\|_\infty < \lambda/8$. We have

$$\begin{aligned} \|Y_{j_0}\|_\infty & = \left\| \sum_{j=1}^{j_0} q^{-1} \mathcal{P}_{\Omega_j} Z_{j-1} \right\|_\infty \\ & \leq q^{-1} \sum_{j=1}^{j_0} \|\mathcal{P}_{\Omega_j} Z_{j-1}\|_\infty \\ & \leq q^{-1} \sum_{j=1}^{j_0} \|Z_{j-1}\|_\infty \\ & \leq q^{-1} \sum_{j=1}^{j_0} \epsilon^{j-1} \sqrt{\frac{\mu r}{mn}} \end{aligned}$$

Notice that $q \geq (1 - \rho)/j_0 \geq 4/\log m$ for $\rho < 1/2$,

$$\begin{aligned}\|Y_{j_0}\|_\infty &\leq \frac{\log m}{4(1 - \epsilon)} \sqrt{\frac{\mu r}{mn}} \\ &\leq \lambda/8,\end{aligned}$$

for sufficiently small ϵ and for some numerical constant C_r .

4.3 Proof of Lemma 4

We recall the notation that $\Gamma^\perp = Q^\perp \oplus T$. By Lemma 11, we have that

$$\|\mathcal{P}_{\Gamma^\perp} \mathcal{P}_\Omega\|^2 \leq \frac{\|\mathcal{P}_\Omega \mathcal{P}_{Q^\perp}\|^2 + \|\mathcal{P}_\Omega \mathcal{P}_T\|^2}{1 - \|\mathcal{P}_{Q^\perp} \mathcal{P}_T\|}.$$

Let us assume that m, n are sufficiently large so that the following conditions hold true:

$$\frac{p}{mn} < \frac{5\rho}{4}, \quad (54)$$

$$8 \left(\sqrt{\frac{p}{mn}} + \sqrt{\frac{(m+n)r}{mn}} \right) < \frac{1}{2}, \quad (55)$$

$$\rho^2(1 - \rho) \geq C_0 \cdot \frac{\mu r \log m}{n}, \quad (56)$$

where $C_0 > 0$ is the numerical constant from Lemma 9. We also assume that $\rho < 1/5$. We note that it is possible to satisfy all of the above inequalities under the assumptions on p and r given in Theorem 1, and because ρ is a fixed constant in the interval $(0, 1)$. Using Lemma 11, it is easy to verify that under these assumptions, with high probability, we have that

$$\|\mathcal{P}_{\Gamma^\perp} \mathcal{P}_\Omega\| \leq \eta \sqrt{\rho}, \quad (57)$$

where $\eta > 0$ is a numerical constant.

The basic steps of the proof closely follow that of Lemma 2.9 in [6]. We recognize that using the convergent Neumann series, W^S can be expressed as follows:

$$W^S = \lambda(\mathcal{I} - \mathcal{P}_{\Gamma^\perp}) \mathcal{P}_\Omega \sum_{k \geq 0} (\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp} \mathcal{P}_\Omega)^k [\text{sgn}(S_0)]. \quad (58)$$

As mentioned in Section 1.2, we assume that the signs of the non-zero entries of S_0 are independent, symmetric ± 1 random variables.

4.3.1 Bounding $\|W^S\|$

It is easy to show that

$$\begin{aligned}W^S &= \lambda \text{sgn}(S_0) - \lambda \mathcal{P}_{\Omega^\perp} \mathcal{P}_{\Gamma^\perp} \mathcal{P}_\Omega \sum_{k \geq 0} (\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp} \mathcal{P}_\Omega)^k [\text{sgn}(S_0)] \\ &:= W_1^S - W_2^S\end{aligned}$$

We now show that each of these components have spectral norm smaller than $1/16$ with high probability. This gives us the desired bound on $\|W^S\|$.

For the first term, we can use standard arguments about the norms of random matrices with i.i.d. entries (see [23]) to show that, with high probability,

$$\|\text{sgn}(S_0)\| \leq 4\sqrt{n\rho}.$$

Since $\lambda = m^{-1/2}$, we have that $\|W_1^S\| \leq 4\sqrt{\rho}$ with high probability. Thus, for sufficiently small ρ , we have that $\|W_1^S\| < 1/16$.

We use a discretization argument to bound $\|W_2^S\|$. Let N_m and N_n be $1/2$ -nets for the unit spheres in \mathbb{R}^m and \mathbb{R}^n , respectively. It can be shown that the sizes of N_m and N_n are at most 6^m and 6^n , respectively (see Theorem 4.16 in [19]). Then, we have that

$$\begin{aligned}
\|W_2^S\| &\leq 4 \max_{\mathbf{x} \in N_m, \mathbf{y} \in N_n} \mathbf{x}^* W_2^S \mathbf{y} \\
&= 4 \max_{\mathbf{x} \in N_m, \mathbf{y} \in N_n} \langle \mathbf{x} \mathbf{y}^*, W_2^S \rangle \\
&= 4 \max_{\mathbf{x} \in N_m, \mathbf{y} \in N_n} \left\langle \mathbf{x} \mathbf{y}^*, \lambda \mathcal{P}_{\Omega^\perp} \mathcal{P}_{\Gamma^\perp} \mathcal{P}_\Omega \sum_{k \geq 0} (\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp} \mathcal{P}_\Omega)^k [\text{sgn}(S_0)] \right\rangle \\
&= 4\lambda \max_{\mathbf{x} \in N_m, \mathbf{y} \in N_n} \left\langle \sum_{k \geq 0} (\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp} \mathcal{P}_\Omega)^k \mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp} \mathcal{P}_{\Omega^\perp} [\mathbf{x} \mathbf{y}^*], \text{sgn}(S_0) \right\rangle \\
&= 4\lambda \max_{\mathbf{x} \in N_m, \mathbf{y} \in N_n} \langle H(\mathbf{x}, \mathbf{y}), \text{sgn}(S_0) \rangle.
\end{aligned}$$

For any $(\mathbf{x}, \mathbf{y}) \in N_m \times N_n$, we bound $\|H(\mathbf{x}, \mathbf{y})\|_F$ as follows:

$$\begin{aligned}
\|H(\mathbf{x}, \mathbf{y})\|_F &= \left\| \sum_{k \geq 0} (\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp} \mathcal{P}_\Omega)^k \mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp} \mathcal{P}_{\Omega^\perp} [\mathbf{x} \mathbf{y}^*] \right\|_F \\
&\leq \sum_{k \geq 0} \|(\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp} \mathcal{P}_\Omega)^k \mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp} \mathcal{P}_{\Omega^\perp} [\mathbf{x} \mathbf{y}^*]\|_F \\
&\leq \left(\sum_{k \geq 0} \|(\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp} \mathcal{P}_\Omega)^k\| \right) \|\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp}\| \|\mathcal{P}_{\Omega^\perp} [\mathbf{x} \mathbf{y}^*]\|_F \\
&\leq \frac{\|\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp}\|}{1 - \|\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp}\|^2}.
\end{aligned}$$

Conditioned on Q and Ω , we use Hoeffding's inequality to get

$$\mathbb{P} [|\langle H(\mathbf{x}, \mathbf{y}), \text{sgn}(S_0) \rangle| > t | \Omega, Q] < 2 \exp \left(-\frac{2t^2}{\|H(\mathbf{x}, \mathbf{y})\|_F^2} \right).$$

Subsequently, using a union bound over $N_m \times N_n$, we obtain

$$\mathbb{P} \left[\max_{\mathbf{x} \in N_m, \mathbf{y} \in N_n} |\langle H(\mathbf{x}, \mathbf{y}), \text{sgn}(S_0) \rangle| > t | \Omega, Q \right] \tag{59}$$

$$< 2 \cdot 6^{m+n} \cdot \exp \left(-\frac{2t^2}{\max_{\mathbf{x} \in N_m, \mathbf{y} \in N_n} \|H(\mathbf{x}, \mathbf{y})\|_F^2} \right) \tag{60}$$

$$\leq 2 \cdot 6^{m+n} \cdot \exp \left(-\frac{2t^2(1 - \|\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp}\|^2)^2}{\|\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp}\|^2} \right). \tag{61}$$

Let E_1 be the event $\{\|\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp}\| \leq \eta\sqrt{\rho}\}$. We know that this event occurs with high probability. Thus, removing the conditioning on Ω and Q , we have

$$\begin{aligned}
\mathbb{P} \left[\max_{\mathbf{x} \in N_m, \mathbf{y} \in N_n} |\langle H(\mathbf{x}, \mathbf{y}), \text{sgn}(S_0) \rangle| > t \right] &< 2 \cdot 6^{m+n} \cdot \exp \left(-\frac{2t^2(1 - \eta^2\rho)^2}{\eta^2\rho} \right) \\
&\quad + \mathbb{P}[E_1^c].
\end{aligned}$$

Therefore,

$$\mathbb{P} \left[4\lambda \max_{\mathbf{x} \in N_m, \mathbf{y} \in N_n} |\langle H(\mathbf{x}, \mathbf{y}), \text{sgn}(S_0) \rangle| > t \right] < 2 \cdot 6^{m+n} \cdot \exp \left(-\frac{t^2(1-\eta^2\rho)^2}{8\lambda^2\eta^2\rho} \right) + \mathbb{P}[E_1^c].$$

Setting $t = \frac{s\eta\sqrt{16\rho}}{1-\eta^2\rho}$ and substituting $\lambda = 1/\sqrt{m}$, we get

$$\mathbb{P} \left[\|W_2^S\| > \frac{s\eta\sqrt{16\rho}}{1-\eta^2\rho} \right] < 2 \cdot \exp(2m(\log 6 - s^2)) + \mathbb{P}[E_1^c].$$

Let us choose any $s > \sqrt{\log 6}$. Then, for sufficiently small ρ , we have that $\|W_2^S\| < 1/16$ with high probability.

4.3.2 Bounding $\|\mathcal{P}_{\Omega^\perp} W^S\|_\infty$

Once again, using the convergent Neumann series expansion for W^S , we have

$$\begin{aligned} \|\mathcal{P}_{\Omega^\perp} W^S\|_\infty &= \max_{(i,j) \in \Omega^c} \left| \left\langle \bar{e}_i e_j^*, \lambda(\mathcal{I} - \mathcal{P}_{\Gamma^\perp}) \mathcal{P}_\Omega \sum_{k \geq 0} (\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp} \mathcal{P}_\Omega)^k [\text{sgn}(S_0)] \right\rangle \right| \\ &= \lambda \max_{(i,j) \in \Omega^c} \left| \left\langle \sum_{k \geq 0} (\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp} \mathcal{P}_\Omega)^k \mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp} [\bar{e}_i e_j^*], \text{sgn}(S_0) \right\rangle \right| \\ &= \lambda \max_{(i,j) \in \Omega^c} |\langle X_{i,j}, \text{sgn}(S_0) \rangle|. \end{aligned}$$

Conditioned on Q and Ω , we use Hoeffding's inequality to get

$$\mathbb{P} [|\langle X_{i,j}, \text{sgn}(S_0) \rangle| > t | \Omega, Q] < 2 \exp \left(-\frac{2t^2}{\|X_{i,j}\|_F^2} \right).$$

Using a union bound, we obtain

$$\mathbb{P} \left[\max_{i,j} |\langle X_{i,j}, \text{sgn}(S_0) \rangle| > t | \Omega, Q \right] < 2mn \exp \left(-\frac{2t^2}{\max_{i,j} \|X_{i,j}\|_F^2} \right).$$

We obtain a bound on $\|X_{i,j}\|_F$ as follows:

$$\begin{aligned} \|X_{i,j}\|_F &= \left\| \sum_{k \geq 0} (\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp} \mathcal{P}_\Omega)^k \mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp} [\bar{e}_i e_j^*] \right\|_F \\ &\leq \sum_{k \geq 0} \|(\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp} \mathcal{P}_\Omega)^k \mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp} [\bar{e}_i e_j^*]\|_F \\ &\leq \left(\sum_{k \geq 0} \|\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp} \mathcal{P}_\Omega\|^k \right) \|\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp} [\bar{e}_i e_j^*]\|_F \\ &\leq \frac{\|\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp}\| \|\mathcal{P}_{\Gamma^\perp} [\bar{e}_i e_j^*]\|_F}{1 - \|\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp}\|^2}. \end{aligned}$$

Thus, we get

$$\begin{aligned} &\mathbb{P} \left[\max_{i,j} |\langle X_{i,j}, \text{sgn}(S_0) \rangle| > t | \Omega, Q \right] \\ &< 2mn \exp \left(-\frac{2t^2(1 - \|\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp}\|^2)^2}{\|\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp}\|^2 \max_{i,j} \|\mathcal{P}_{\Gamma^\perp} [\bar{e}_i e_j^*]\|_F^2} \right). \end{aligned}$$

Removing the conditioning on Q and Ω , we get

$$\mathbb{P} \left[\|\mathcal{P}_{\Omega^\perp} W^S\|_\infty > \lambda \sqrt{\frac{s \log(mn)}{2}} \frac{\|\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp}\| \max_{i,j} \|\mathcal{P}_{\Gamma^\perp} [\bar{e}_i e_j^*]\|_F}{1 - \|\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp}\|^2} \right] < 2(mn)^{1-s}$$

Consider the two events:

$$\begin{aligned} E_1 &:= \{ \|\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp}\| \leq \eta \sqrt{\rho} \}, \\ E_2 &:= \left\{ \max_{i,j} \|\mathcal{P}_{\Gamma^\perp} \bar{e}_i e_j^*\|_F \leq \sqrt{\gamma} \right\}, \end{aligned}$$

where we recall that $\gamma = 4 \left(\frac{8p \log(mnp)}{mn} + \frac{\mu r}{n} \right)$. We have already shown that E_1 and E_2 occur with high probability. Substituting for the various bounds and setting $s = 2$, we get

$$\begin{aligned} &\mathbb{P} \left[\|\mathcal{P}_{\Omega^\perp} W^S\|_\infty > \lambda \sqrt{\gamma \log(mn)} \frac{\eta \sqrt{\rho}}{1 - \eta^2 \rho} \right] \\ &< \frac{2}{mn} + \mathbb{P}[(E_1 \cap E_2)^c]. \end{aligned}$$

Under the conditions of Theorem 1, and for sufficiently large m, n and sufficiently small ρ , we get that $\|\mathcal{P}_{\Omega^\perp} W^S\|_\infty < \lambda/8$ with high probability.

4.4 Proof of Lemma 5

4.4.1 Bounding $\|W^Q\|$

Using the convergent Neumann series expansion, we can write the analytical expression for W^Q as follows:

$$W^Q = \mathcal{P}_{\Pi^\perp} \sum_{k \geq 0} (\mathcal{P}_{Q^\perp} \mathcal{P}_\Pi \mathcal{P}_{Q^\perp})^k (\mathcal{P}_{Q^\perp} (-UV^*)), \quad (62)$$

where we recall that $\Pi = \Omega \oplus T$. It follows that

$$\|W^Q\|_F \leq \left\| \sum_{k \geq 0} (\mathcal{P}_{Q^\perp} \mathcal{P}_\Pi \mathcal{P}_{Q^\perp})^k \right\| \|\mathcal{P}_{Q^\perp} (UV^*)\|_F.$$

Considering the first term of the product on the right hand side,

$$\begin{aligned} \left\| \sum_{k \geq 0} (\mathcal{P}_{Q^\perp} \mathcal{P}_\Pi \mathcal{P}_{Q^\perp})^k \right\| &\leq \sum_{k \geq 0} \|(\mathcal{P}_{Q^\perp} \mathcal{P}_\Pi \mathcal{P}_{Q^\perp})^k\| \\ &\leq \sum_{k \geq 0} \|\mathcal{P}_{Q^\perp} \mathcal{P}_\Pi\|^{2k}. \end{aligned}$$

From Lemma 11, we have that, for any $\epsilon > 0$, with high probability,

$$\begin{aligned} &\|\mathcal{P}_{Q^\perp} \mathcal{P}_\Pi\|^2 \\ &\leq \frac{64}{1 - \sqrt{\rho} + \epsilon} \left(\left(\sqrt{\frac{p}{mn}} + \sqrt{\frac{5\rho}{4}} \right)^2 + \left(\sqrt{\frac{p}{mn}} + \sqrt{\frac{(m+n)r}{mn}} \right)^2 \right). \end{aligned}$$

Assume that $\rho < 1/4$, and fix $\epsilon = 3\rho$. For m, n large enough, we can assume that $\max\{p/mn, r(m+n)/mn\} < \rho$. Then, we have that, with high probability,

$$\|\mathcal{P}_{Q^\perp} \mathcal{P}_\Pi\|^2 \leq \frac{832\rho}{1 - 2\sqrt{\rho}}.$$

Therefore, for sufficiently small ρ , we have that

$$\|\mathcal{P}_{Q^\perp} \mathcal{P}_\Pi\|^2 \leq \frac{1}{4}, \quad (63)$$

with high probability. Consequently,

$$\left\| \sum_{k \geq 0} (\mathcal{P}_{Q^\perp} \mathcal{P}_\Pi \mathcal{P}_{Q^\perp})^k \right\| \leq \frac{4}{3}, \quad (64)$$

with high probability.

We bound $\|\mathcal{P}_{Q^\perp}(UV^*)\|_F$ as follows. As explained earlier, suppose we vectorize all matrices, then \mathcal{P}_{Q^\perp} has the same distribution as $H(H^*H)^{-1}H^*$, where $H \in \mathbb{R}^{mn \times p}$ is a random Gaussian matrix with i.i.d. entries $\sim \mathcal{N}(0, 1/mn)$. Therefore, we have

$$\|\mathcal{P}_{Q^\perp}(UV^*)\|_F = \|H(H^*H)^{-1}H^*\text{vec}(UV^*)\|_2 \leq \|H(H^*H)^{-1}\| \|H^*\text{vec}(UV^*)\|_2,$$

where the above equality is in distribution. We have already shown in the proof of Lemma 7 that

$$\mathbb{P} [\|H(H^*H)^{-1}\| \geq 4] \leq e^{-mn/32}.$$

We note that $H^*\text{vec}(UV^*)$ is a p -dimensional vector whose components are i.i.d. and have the same distribution as $\langle G, UV^* \rangle$, where $G \in \mathbb{R}^{m \times n}$ is a random Gaussian matrix whose entries are i.i.d. $\sim \mathcal{N}(0, 1/mn)$. It is easy to see that $\langle G, UV^* \rangle$ is distributed according to $\mathcal{N}(0, r/mn)$, and therefore, we have

$$\mathbb{E}[\|H^*\text{vec}(UV^*)\|_F] \leq (\mathbb{E}[\|H^*\text{vec}(UV^*)\|_F^2])^{1/2} = \sqrt{\frac{pr}{mn}}.$$

Since $\|\cdot\|_F$ is a 1-Lipschitz function, we use Proposition 2.18 in [19] to get

$$\mathbb{P} \left[\|H^*\text{vec}(UV^*)\|_F \geq \mathbb{E}[\|H^*\text{vec}(UV^*)\|_F] + t \cdot \sqrt{\frac{r}{mn}} \right] \leq e^{-t^2/2}.$$

Setting $t = \sqrt{6 \log m}$, we get

$$\mathbb{P} \left(\|H^*\text{vec}(UV^*)\|_F \geq \sqrt{\frac{pr}{mn}} + \sqrt{\frac{6r \log m}{mn}} \right) \leq \frac{1}{m^3}. \quad (65)$$

Putting it all together, we conclude that

$$\|W^Q\| \leq \|W^Q\|_F \leq \frac{16}{3} \left(\sqrt{\frac{pr}{mn}} + \sqrt{\frac{6r \log m}{mn}} \right), \quad (66)$$

with high probability. Clearly, for sufficiently large m , the right hand side can be made arbitrarily small under the conditions of Theorem 1 and hence, we have the desired bound.

4.4.2 Controlling $\|\mathcal{P}_{\Omega^\perp} W^Q\|_\infty$

It is easy to show that the analytical expression for W^Q can be written slightly differently as follows:

$$W^Q = \mathcal{P}_{\Pi^\perp} \mathcal{P}_{Q^\perp} \sum_{k \geq 0} (\mathcal{P}_{Q^\perp} \mathcal{P}_\Pi \mathcal{P}_{Q^\perp})^k (\mathcal{P}_{Q^\perp}(-UV^*)). \quad (67)$$

Consider any $(i, j) \in [m] \times [n]$. Then,

$$\begin{aligned}
|\langle W^Q, \bar{e}_i e_j^* \rangle| &= \left| \left\langle \sum_{k \geq 0} (\mathcal{P}_{Q^\perp} \mathcal{P}_\Pi \mathcal{P}_{Q^\perp})^k (\mathcal{P}_{Q^\perp} (-UV^*)), \mathcal{P}_{Q^\perp} \mathcal{P}_{\Pi^\perp} \bar{e}_i e_j^* \right\rangle \right| \\
&\leq \left\| \sum_{k \geq 0} (\mathcal{P}_{Q^\perp} \mathcal{P}_\Pi \mathcal{P}_{Q^\perp})^k \mathcal{P}_{Q^\perp} (UV^*) \right\|_F \|\mathcal{P}_{Q^\perp} \mathcal{P}_{\Pi^\perp} \bar{e}_i e_j^*\|_F \\
&\leq \left\| \sum_{k \geq 0} (\mathcal{P}_{Q^\perp} \mathcal{P}_\Pi \mathcal{P}_{Q^\perp})^k \right\| \|\mathcal{P}_{Q^\perp} (UV^*)\|_F \|\mathcal{P}_{Q^\perp} \mathcal{P}_{\Pi^\perp} \bar{e}_i e_j^*\|_F.
\end{aligned}$$

We have already derived bounds for the first two terms. For the final term in the product, we use the same technique we employed to bound $\|\mathcal{P}_{Q^\perp} (UV^*)\|_F$. Using the fact that $\|\mathcal{P}_{\Pi^\perp} \bar{e}_i e_j^*\|_F \leq 1$, we can show that

$$\mathbb{P} \left[\|\mathcal{P}_{Q^\perp} \mathcal{P}_{\Pi^\perp} \bar{e}_i e_j^*\|_F > \frac{16}{3} \left(\sqrt{\frac{p}{mn}} + \sqrt{\frac{6 \log m}{mn}} \right) \right] \leq \frac{1}{m^3} + e^{-mn/32}.$$

Using a union bound, we get

$$\mathbb{P} \left[\max_{i,j} \|\mathcal{P}_{Q^\perp} \mathcal{P}_{\Pi^\perp} \bar{e}_i e_j^*\|_F > \frac{16}{3} \left(\sqrt{\frac{p}{mn}} + \sqrt{\frac{6 \log m}{mn}} \right) \right] \leq mn(m^{-3} + e^{-mn/32}).$$

Putting all the bounds together, we have that, with high probability,

$$\|\mathcal{P}_{Q^\perp} W^Q\|_\infty \leq \frac{256}{9} \frac{\sqrt{r}}{mn} \left(\sqrt{p} + \sqrt{6 \log m} \right)^2. \quad (68)$$

Since $\lambda = m^{-1/2}$, it is easy to show that under the assumptions of Theorem 1, the right hand side in the above inequality can be made smaller than $C' \lambda$, for any fixed $C' > 0$. Thus, we have the desired bound.

5 Deterministic Reduction: Proof of Theorem 2

In this section, we provide the proof for Theorem 2 under the deterministic subspace model for Q^\perp . We will adopt the same optimality conditions established in Lemma 2, and the same proof strategy outlined in Section 3, namely the construction of $W = W^L + W^S + W^Q$. To avoid redundancy, wherever possible, we will only highlight the parts that differ from the previous proof in Section 4 and refer the interested reader to Section 4 for more details. First, we derive the various incoherence relations associated with our fixed subspace Q^\perp . Then, we will prove Lemmas 3, 4 and 5 using these relations.

5.1 Preliminaries

In this subsection, we provide several lemmas that will be used later in our proof.

Lemma 15. If $X \in \mathbb{R}^{m \times n}$ is a rank- r matrix, then

$$\|\mathcal{P}_{Q^\perp} X\|_F^2 \leq \nu \frac{pr}{n} \|X\|_F^2. \quad (69)$$

Proof.

$$\begin{aligned}
\|\mathcal{P}_{Q^\perp} X\|_F^2 &= \sum_{i=1}^p |\langle G_i, X \rangle|^2 \\
&\leq p \left(\max_i \|G_i\|^2 \right) \|X\|_*^2 \\
&\leq pr \left(\max_i \|G_i\|^2 \right) \|X\|_F^2 \\
&\leq \nu \frac{pr}{n} \|X\|_F^2.
\end{aligned}$$

□

Corollary 1. For any ν -coherent subspace Q^\perp , we have the following:

1. $\|\mathcal{P}_{Q^\perp} \bar{e}_i e_j^*\|_F^2 \leq \nu \frac{p}{n}$;
2. $\|\mathcal{P}_{Q^\perp} \mathcal{P}_T\|^2 \leq 2\nu \frac{pr}{n}$;
3. $\|\mathcal{P}_{Q^\perp} (UV^*)\|_F^2 \leq 2\nu \frac{pr^2}{n}$.

Proof. The first two results follow from the fact that the $\bar{e}_i e_j^*$ are rank-1 matrices, and $\text{rank}(\mathcal{P}_T X) \leq 2r \forall X \in \mathbb{R}^{m \times n}$. The last result can be derived from the second one as shown below:

$$\|\mathcal{P}_{Q^\perp} (UV^*)\|_F^2 \leq \|\mathcal{P}_{Q^\perp} \mathcal{P}_T\|^2 \|UV^*\|_F^2 \leq 2\nu \frac{pr^2}{n}.$$

□

Lemma 16. Under the assumptions made in Theorem 2, we have that

$$\|\mathcal{P}_{Q^\perp} \mathcal{P}_\Omega\| < 1/2, \quad (70)$$

with high probability, provided that $\rho < \rho_0$ and $\nu^2 p^3 \log m/n \leq C$. Here, $C > 0$ and $\rho_0 \in (0, 1)$ are numerical constants.

Proof. Please refer to Section 5.5 for a detailed proof. □

5.2 Proof of Lemma 3 (deterministic case)

We use the same framework from Section 4.2.1 to bound the corresponding norms of W^L . We note that to bound $\|W^L\|$ in the previous case, the only key property of Q^\perp that was critical to the proof was that $\Gamma^\perp = Q^\perp \oplus T$ is $O(\mu r/n)$ -constrained. More specifically, the latter property is used in Lemma 12 and Lemma 13.

In the deterministic case, by assumption, Q^\perp is ν -coherent, where ν is a constant. In the following lemma, we will show that Γ^\perp is $O(\mu r/n)$ -constrained as well under our assumptions. We will show that, the proof of Lemma 3 can be directly adopted for the deterministic case from the that with the random subspace model.

Lemma 17. If Q^\perp is ν -coherent, then

$$\|\mathcal{P}_{\Gamma^\perp} \bar{e}_i e_j^*\|_F \leq 4 \left(\sqrt{\frac{\nu p}{n}} + \sqrt{\frac{2\mu r}{n}} \right). \quad (71)$$

In other words, if Q^\perp is ν -coherent, then Γ^\perp is γ -constrained for $\gamma = 16 \left(\sqrt{\nu p/n} + \sqrt{2\mu r/n} \right)^2$.

Proof. Let us assume that $\|\mathcal{P}_{Q^\perp}\mathcal{P}_T\| < 1/2$. This is true for sufficiently large n under the assumptions of Theorem 2. Using the convergent Neumann series expansion, it is possible to show that

$$\mathcal{P}_{\Gamma^\perp}\bar{e}_i e_j^* = ((\mathcal{I} - \mathcal{P}_{Q^\perp}\mathcal{P}_T)^{-1}\mathcal{P}_{Q^\perp}\mathcal{P}_{T^\perp} + (\mathcal{I} - \mathcal{P}_T\mathcal{P}_{Q^\perp})^{-1}\mathcal{P}_T\mathcal{P}_Q)(\bar{e}_i e_j^*),$$

and therefore,

$$\|\mathcal{P}_{\Gamma^\perp}\bar{e}_i e_j^*\|_F \leq \|\mathcal{I} - \mathcal{P}_{Q^\perp}\mathcal{P}_T\|^{-1} \|\mathcal{P}_{Q^\perp}\mathcal{P}_{T^\perp}(\bar{e}_i e_j^*)\|_F + \|\mathcal{I} - \mathcal{P}_T\mathcal{P}_{Q^\perp}\|^{-1} \|\mathcal{P}_T\mathcal{P}_Q(\bar{e}_i e_j^*)\|_F.$$

From Eqn. (6) and Corollary 1, we have

$$\begin{aligned} \|\mathcal{P}_{Q^\perp}\mathcal{P}_{T^\perp}(\bar{e}_i e_j^*)\|_F &\leq \|\mathcal{P}_{Q^\perp}(\bar{e}_i e_j^*)\|_F + \|\mathcal{P}_{Q^\perp}\mathcal{P}_T(\bar{e}_i e_j^*)\|_F \\ &\leq \|\mathcal{P}_{Q^\perp}(\bar{e}_i e_j^*)\|_F + \|\mathcal{P}_T(\bar{e}_i e_j^*)\|_F \\ &\leq \sqrt{\frac{\nu p}{n}} + \sqrt{\frac{2\mu r}{n}}. \end{aligned}$$

Similarly, we have

$$\|\mathcal{P}_T\mathcal{P}_Q(\bar{e}_i e_j^*)\|_F \leq \sqrt{\frac{\nu p}{n}} + \sqrt{\frac{2\mu r}{n}}.$$

We also have that

$$\|(\mathcal{I} - \mathcal{P}_{Q^\perp}\mathcal{P}_T)^{-1}\| = \|(\mathcal{I} - \mathcal{P}_T\mathcal{P}_{Q^\perp})^{-1}\| = \left\| \sum_{k \geq 0} (\mathcal{P}_{Q^\perp}\mathcal{P}_T)^k \right\| < 2.$$

Hence, we have

$$\|\mathcal{P}_{\Gamma^\perp} e_i e_j^*\|_F \leq 4 \left(\sqrt{\frac{\nu p}{n}} + \sqrt{\frac{2\mu r}{n}} \right). \quad (72)$$

□

It can be easily shown that the results in Section 4.1 all hold for the deterministic case as well with the modified value for γ derived above. Consequently, the proof of Lemma 3 from Section 4.2 can be directly adopted for the deterministic case as well.

5.3 Proof of Lemma 4 (deterministic case)

We now provide a proof of Lemma 4 under our deterministic subspace model. Since the basic framework of the proof is very similar to that in Section 4.3, we will derive only the important steps here and refer the interested reader to Section 4.3 for more details.

Controlling $\|\mathcal{P}_{\Omega^\perp} W^S\|_\infty$ Using the convergent Neumann series, we have

$$W^S = \lambda (\mathcal{I} - \mathcal{P}_{\Gamma^\perp}) \mathcal{P}_\Omega \sum_{k \geq 0} (\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp} \mathcal{P}_\Omega)^k [\text{sgn}(S_0)].$$

Therefore, we have

$$\begin{aligned} \|\mathcal{P}_{\Omega^\perp} W^S\|_\infty &= \lambda \max_{(i,j) \in \Omega^c} \left| \left\langle \bar{e}_i e_j^*, (\mathcal{I} - \mathcal{P}_{\Gamma^\perp}) \mathcal{P}_\Omega \sum_{k \geq 0} (\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp} \mathcal{P}_\Omega)^k [\text{sgn}(S_0)] \right\rangle \right| \\ &= \lambda \max_{(i,j) \in \Omega^c} \left| \left\langle \sum_{k \geq 0} (\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp} \mathcal{P}_\Omega)^k \mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp}(\bar{e}_i e_j^*), \text{sgn}(S_0) \right\rangle \right| \\ &= \lambda \max_{(i,j) \in \Omega^c} \left| \left\langle H^{(i,j)}, \text{sgn}(S_0) \right\rangle \right|. \end{aligned}$$

We now bound $\|H^{(i,j)}\|_F$ as follows:

$$\begin{aligned}
\|H^{(i,j)}\|_F &\leq \sum_{k \geq 0} \|(\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp} \mathcal{P}_\Omega)^k \mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp} (\bar{e}_i e_j^*)\|_F \\
&\leq \left(\sum_{k \geq 0} \|(\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp} \mathcal{P}_\Omega)^k\| \right) \|\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp}\| \|\mathcal{P}_{\Gamma^\perp} (\bar{e}_i e_j^*)\|_F \\
&\leq \frac{\|\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp}\| \|\mathcal{P}_{\Gamma^\perp} (\bar{e}_i e_j^*)\|_F}{1 - \|\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp}\|^2}.
\end{aligned}$$

Conditioned on Ω , using Hoeffding's inequality, we have

$$\mathbb{P} \left[\left| \langle H^{(i,j)}, \text{sgn}(S_0) \rangle \right| > t \mid \Omega \right] < 2 \exp \left(-\frac{2t^2}{\|H^{(i,j)}\|_F^2} \right).$$

Applying a union bound, we get

$$\begin{aligned}
\mathbb{P} \left[\max_{i,j} \left| \langle H^{(i,j)}, \text{sgn}(S_0) \rangle \right| > t \mid \Omega \right] &\leq 2mn \exp \left(-\frac{2t^2}{\max_{i,j} \|H^{(i,j)}\|_F^2} \right) \\
&\leq 2mn \exp \left(-\frac{2t^2 (1 - \|\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp}\|^2)^2}{\|\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp}\|^2 \max_{i,j} \|\mathcal{P}_{\Gamma^\perp} (\bar{e}_i e_j^*)\|_F^2} \right).
\end{aligned}$$

Removing the conditioning on Ω , we get

$$\mathbb{P} \left[\|\mathcal{P}_{\Omega^\perp} W^S\|_\infty > \lambda \sqrt{\frac{s \log(mn)}{2}} \frac{\|\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp}\| \max_{i,j} \|\mathcal{P}_{\Gamma^\perp} (\bar{e}_i e_j^*)\|_F}{1 - \|\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp}\|^2} \right] < 2(mn)^{1-s},$$

where $s > 0$. Consider the event $E = \{\|\mathcal{P}_\Omega \mathcal{P}_{\Gamma^\perp}\| \leq \eta \sqrt{\rho}\}$. Just like under the random subspace model, it is not difficult to show that the event E occurs with high probability for some fixed $\eta > 0$. Furthermore, we have already shown that Γ^\perp is a γ -constrained subspace with $\gamma \log m = O(1/\log m)$. Setting $s = 2$, we get

$$\mathbb{P} \left[\|\mathcal{P}_{\Omega^\perp} W^S\|_\infty > \lambda \sqrt{\gamma \log(mn)} \frac{\eta \sqrt{\rho}}{1 - \eta^2 \rho} \right] < \frac{2}{mn} + \mathbb{P}[E^c].$$

Thus, we have the desired bound.

Controlling $\|W^S\|$ The proof is identical to the one in Section 4.3.1.

5.4 Proof of Lemma 5 (deterministic case)

We now prove Lemma 5 under our deterministic subspace model. Once again, the basic structure of the proof is very similar to the one used in Section 4.4. So, we only provide the relevant bounds here and refer the interested reader to Section 4.4 for the detailed steps involved.

Controlling $\|W^Q\|$ The proof framework is the same as the one in Section 4.4.1. We note that the key step is to bound $\|\mathcal{P}_{Q^\perp}(UV^*)\|_F$ and $\left\| \sum_{k \geq 0} (\mathcal{P}_{Q^\perp} \mathcal{P}_\Pi \mathcal{P}_{Q^\perp})^k \right\|$, where we recall that $\Pi = \Omega \oplus T$. From Corollary 1, we already know that

$$\|\mathcal{P}_{Q^\perp}(UV^*)\|_F \leq \sqrt{\frac{2\nu p r^2}{n}}.$$

For the other quantity, we have that

$$\left\| \sum_{k \geq 0} (\mathcal{P}_{Q^\perp} \mathcal{P}_\Pi \mathcal{P}_{Q^\perp})^k \right\| \leq \frac{1}{1 - \|\mathcal{P}_{Q^\perp} \mathcal{P}_\Pi\|^2}.$$

By Lemma 11, we have

$$\|\mathcal{P}_{Q^\perp} \mathcal{P}_\Pi\|^2 \leq \frac{\|\mathcal{P}_{Q^\perp} \mathcal{P}_\Omega\|^2 + \|\mathcal{P}_{Q^\perp} \mathcal{P}_T\|^2}{1 - \|\mathcal{P}_\Omega \mathcal{P}_T\|}.$$

From Lemma 9, we know that $\|\mathcal{P}_\Omega \mathcal{P}_T\| \leq \sqrt{\rho} + \epsilon$ with high probability, provided that

$$(1 - \rho) \geq C_0 \cdot \epsilon^{-2} \frac{\mu r \log m}{n}.$$

Suppose that the above condition holds with $\epsilon = \rho$, and assume that

$$\frac{2\nu pr}{n} < \rho.$$

We note that both the assumptions above can be true for sufficiently large m and n under the assumptions of Theorem 2. Under these assumptions, along with Lemma 16, we have

$$\|\mathcal{P}_{Q^\perp} \mathcal{P}_\Pi\|^2 \leq \frac{1/4 + \rho}{1 - \sqrt{2\rho}},$$

with high probability. Thus, we have that $\|\mathcal{P}_{Q^\perp} \mathcal{P}_\Pi\|^2 \leq 1/2$ with high probability, provided that ρ is sufficiently small. Putting all these bounds together, we get

$$\|W^Q\| \leq 2\sqrt{\frac{2\nu pr^2}{n}},$$

with high probability. Under the assumptions of Theorem 2, the right hand side can be made arbitrarily small, and hence, we have the desired result.

Controlling $\|\mathcal{P}_{\Omega^\perp} W^Q\|_\infty$ Once again, the proof framework is identical to that used in Section 4.4.2. The key step here is to bound $\max_{(i,j) \in \Omega^c} \|\mathcal{P}_{Q^\perp} \mathcal{P}_{\Pi^\perp} \bar{e}_i e_j^*\|_F$. We first use the Neumann series to rewrite $\mathcal{P}_\Pi \bar{e}_i e_j^*$ as

$$\mathcal{P}_\Pi \bar{e}_i e_j^* = ((\mathcal{I} - \mathcal{P}_\Omega \mathcal{P}_T)^{-1} \mathcal{P}_\Omega \mathcal{P}_{T^\perp} + (\mathcal{I} - \mathcal{P}_T \mathcal{P}_\Omega)^{-1} \mathcal{P}_T \mathcal{P}_{\Omega^\perp}) (\bar{e}_i e_j^*).$$

Now, for any $(i, j) \in [m] \times [n]$, we have

$$\begin{aligned} \|\mathcal{P}_\Omega \mathcal{P}_{T^\perp} \bar{e}_i e_j^*\|_F = \|\mathcal{P}_\Omega \mathcal{P}_T \bar{e}_i e_j^*\|_F &\leq \sqrt{\frac{2\mu r}{n}}, \\ \|\mathcal{P}_T \mathcal{P}_{\Omega^\perp} \bar{e}_i e_j^*\|_F = \|\mathcal{P}_T \bar{e}_i e_j^*\|_F &\leq \sqrt{\frac{2\mu r}{n}}. \end{aligned}$$

Furthermore, by the assumption we used earlier (to bound $\|W^Q\|$), we have that $\|\mathcal{P}_\Omega \mathcal{P}_T\| < \sqrt{2\rho}$ with high probability. Therefore, we have

$$\begin{aligned} \|(\mathcal{I} - \mathcal{P}_\Omega \mathcal{P}_T)^{-1}\| &= \|(\mathcal{I} - \mathcal{P}_T \mathcal{P}_\Omega)^{-1}\| \\ &= \left\| \sum_{k \geq 0} (\mathcal{P}_\Omega \mathcal{P}_T)^k \right\| \\ &\leq \frac{1}{1 - \sqrt{2\rho}} \\ &\leq 2 \end{aligned}$$

with high probability, provided that $\rho \leq 1/8$. Thus, we get

$$\|\mathcal{P}_\Pi \bar{e}_i e_j^*\|_F \leq 4\sqrt{\frac{2\mu r}{n}},$$

with high probability. Consequently, for any $(i, j) \in \Omega^c$, we have

$$\begin{aligned} \|\mathcal{P}_{Q^\perp} \mathcal{P}_{\Pi^\perp} \bar{e}_i e_j^*\|_F &\leq \|\mathcal{P}_{Q^\perp} \bar{e}_i e_j^*\|_F + \|\mathcal{P}_{Q^\perp} \mathcal{P}_\Pi \bar{e}_i e_j^*\|_F \\ &\leq \sqrt{\frac{\nu p}{n}} + 4\sqrt{\frac{2\mu r}{n}}, \end{aligned}$$

with high probability.

Proceeding along the same lines as in Section 4.4.2, we have that

$$|\langle W^Q, \bar{e}_i e_j^* \rangle| \leq 2\sqrt{\frac{2\nu p r^2}{n}} \left(\sqrt{\frac{\nu p}{n}} + 4\sqrt{\frac{2\mu r}{n}} \right),$$

with high probability, for any $(i, j) \in \Omega^c$. Therefore, we have that

$$\|\mathcal{P}_{\Omega^\perp} W^Q\|_\infty \leq 2\sqrt{\frac{2\nu p r^2}{n}} \left(\sqrt{\frac{\nu p}{n}} + 4\sqrt{\frac{2\mu r}{n}} \right),$$

with high probability. Since $\lambda = m^{-1/2}$, under the assumptions made in Theorem 2, the right hand side can be made smaller than $\lambda/8$, provided that n is sufficiently large.

5.5 Proof of Lemma 16

Consider the linear operator

$$\mathcal{A} = \mathcal{P}_{Q^\perp} \mathcal{P}_\Omega \mathcal{P}_{Q^\perp} - \rho \mathcal{P}_{Q^\perp}.$$

It can be easily shown that

$$\mathbb{E}[\mathcal{A}] = 0.$$

First, we derive a bound for the spectral norm of \mathcal{A} . Let δ_{ij} be a sequence of independent Bernoulli random variables such that

$$\delta_{ij} = \begin{cases} 1, & \text{if } (i, j) \in \Omega, \\ 0, & \text{otherwise.} \end{cases}$$

Then, we can rewrite \mathcal{A} as

$$\mathcal{A} = \sum_{ij} \mathcal{A}_{ij},$$

where

$$\mathcal{A}_{ij} = \delta_{ij} \mathcal{P}_{Q^\perp} (\bar{e}_i e_j^*) \otimes \mathcal{P}_{Q^\perp} (\bar{e}_i e_j^*) - \frac{\rho}{mn} \mathcal{P}_{Q^\perp},$$

and \otimes denotes the outer or tensor product between matrices. Then, we have that

$$\|\mathcal{A}_{ij}\| \leq \|\mathcal{P}_{Q^\perp} (\bar{e}_i e_j^*) \otimes \mathcal{P}_{Q^\perp} (\bar{e}_i e_j^*)\| + \frac{\rho}{mn} \quad (73)$$

$$\leq \|\mathcal{P}_{Q^\perp} (\bar{e}_i e_j^*)\|_F^2 + \frac{\rho}{mn} \quad (74)$$

$$\leq \frac{\nu p}{n} + \frac{\rho}{mn} \quad (75)$$

$$\triangleq S, \quad (76)$$

where in Eqn. (74) we used the fact that $\|A \otimes B\| \leq \|A\|_F \|B\|_F$.

We now bound the variance terms.

$$\sigma^2 = \left\| \sum_{i,j} \mathbb{E} [\mathcal{A}_{ij}^2] \right\| \quad (77)$$

$$= \left\| \sum_{i,j} \left(\rho [\mathcal{P}_{Q^\perp}(\bar{e}_i e_j^*) \otimes \mathcal{P}_{Q^\perp}(\bar{e}_i e_j^*)]^2 - \frac{2\rho^2 \mathcal{P}_{Q^\perp}(\bar{e}_i e_j^*) \otimes \mathcal{P}_{Q^\perp}(\bar{e}_i e_j^*)}{mn} + \frac{\rho^2}{m^2 n^2} \mathcal{P}_{Q^\perp} \right) \right\| \quad (78)$$

We let $\mathcal{P}_{\Omega_{ij}}$ denote the orthogonal projector onto the subspace $\text{span}(\bar{e}_i e_j^*)$. Clearly, we have

$$\mathcal{P}_\Omega = \sum_{(i,j) \in \Omega} \mathcal{P}_{\Omega_{ij}}.$$

Furthermore, we note that

$$\mathcal{P}_{Q^\perp}(\bar{e}_i e_j^*) \otimes \mathcal{P}_{Q^\perp}(\bar{e}_i e_j^*) = \mathcal{P}_{Q^\perp} \mathcal{P}_{\Omega_{ij}} \mathcal{P}_{Q^\perp}.$$

Thus, we get

$$\begin{aligned} \sum_{i,j} \frac{2\rho^2 \mathcal{P}_{Q^\perp}(\bar{e}_i e_j^*) \otimes \mathcal{P}_{Q^\perp}(\bar{e}_i e_j^*)}{mn} &= \frac{2\rho^2 \mathcal{P}_{Q^\perp} \left(\sum_{i,j} \mathcal{P}_{\Omega_{ij}} \right) \mathcal{P}_{Q^\perp}}{mn} \\ &= \frac{2\rho^2 \mathcal{P}_{Q^\perp}}{mn}. \end{aligned}$$

Similarly, we have

$$\begin{aligned} \sum_{i,j} \rho [\mathcal{P}_{Q^\perp}(\bar{e}_i e_j^*) \otimes \mathcal{P}_{Q^\perp}(\bar{e}_i e_j^*)]^2 &= \rho \sum_{i,j} \mathcal{P}_{Q^\perp} \mathcal{P}_{\Omega_{ij}} \mathcal{P}_{Q^\perp} \mathcal{P}_{\Omega_{ij}} \mathcal{P}_{Q^\perp} \\ &= \rho \mathcal{P}_{Q^\perp} \left(\sum_{i,j} \mathcal{P}_{\Omega_{ij}} \mathcal{P}_{Q^\perp} \mathcal{P}_{\Omega_{ij}} \right) \mathcal{P}_{Q^\perp}. \end{aligned}$$

Let $X \in \mathbb{R}^{m \times n}$ be any matrix satisfying $\|X\|_F = 1$. Then,

$$\begin{aligned} \left\| \sum_{i,j} \mathcal{P}_{\Omega_{ij}} \mathcal{P}_{Q^\perp} \mathcal{P}_{\Omega_{ij}} X \right\|_F &= \left\| \sum_{i,j} \mathcal{P}_{\Omega_{ij}} \left(\sum_{k=1}^p \langle G_k, \mathcal{P}_{\Omega_{ij}} X \rangle G_k \right) \right\|_F \\ &= \left\| \sum_{i,j} \mathcal{P}_{\Omega_{ij}} \left(\sum_{k=1}^p \langle \mathcal{P}_{\Omega_{ij}} G_k, X \rangle G_k \right) \right\|_F, \end{aligned}$$

where we recall that the G_i 's constitute an orthonormal basis for Q^\perp satisfying $\max_i \|G_i\|^2 < \nu/n$. We now bound $\|\mathcal{P}_{\Omega_{ij}} G_k\|_F$ as follows:

$$\begin{aligned} \|\mathcal{P}_{\Omega_{ij}} G_k\|_F &= |\langle \bar{e}_i e_j^*, \mathcal{P}_{Q^\perp} G_k \rangle| \\ &= |\langle \mathcal{P}_{Q^\perp} \bar{e}_i e_j^*, G_k \rangle| \\ &\leq \|G_k\| \|\mathcal{P}_{Q^\perp} \bar{e}_i e_j^*\|_* \\ &\leq \|G_k\| \sqrt{n} \|\mathcal{P}_{Q^\perp} \bar{e}_i e_j^*\|_F \\ &\leq \sqrt{\frac{\nu}{n}} \sqrt{n} \sqrt{\frac{\nu p}{n}} \\ &= \nu \sqrt{\frac{p}{n}}. \end{aligned}$$

Combining the above bound with Hölder's inequality, we get

$$\left\| \sum_{i,j} \mathcal{P}_{\Omega_{ij}} \mathcal{P}_{Q^\perp} \mathcal{P}_{\Omega_{ij}}(X) \right\|_F \leq \left\| \sum_{i,j,k} \langle \mathcal{P}_{\Omega_{ij}} G_k, X \rangle \mathcal{P}_{\Omega_{ij}} G_k \right\|_F \quad (79)$$

$$\leq \nu \sqrt{\frac{p}{n}} \left\| \left(\sum_{i,j} \mathcal{P}_{\Omega_{ij}} \right) \left(\sum_{k=1}^p G_k \right) \right\|_F \quad (80)$$

$$\leq \nu \sqrt{\frac{p^3}{n}}. \quad (81)$$

Therefore, we have that the variance in Eqn. (78) can be bounded as

$$\begin{aligned} \sigma^2 &\leq \rho \nu \sqrt{\frac{p^3}{n}} + \frac{2\rho^2}{mn} + \frac{\rho^2}{mn} \\ &\leq 2\rho \nu \sqrt{\frac{p^3}{n}} \end{aligned}$$

Applying the matrix Bernstein inequality (Theorem 3), we get

$$\begin{aligned} \mathbb{P} [\|\mathcal{A}\| > t] &\leq 2m^2 \exp \left(-\frac{t^2}{2\sigma^2 + 3St} \right) \\ &\leq 2m^2 \exp \left(-\frac{t^2}{C_1 \rho \nu \sqrt{p^3/m} + C_2 \nu p t / m} \right). \end{aligned}$$

Let us set $t = \rho$. Now, suppose that

$$\frac{\nu^2 p^3 \log m}{n} \leq C_3 \rho^2, \quad (82)$$

where $C_3 > 0$ is a numerical constant. Then, under the conditions of Theorem 2, $\|\mathcal{A}\|$ is bounded from above by ρ with high probability. Since $\mathcal{A} = \mathcal{P}_{Q^\perp} \mathcal{P}_\Omega \mathcal{P}_{Q^\perp} - \rho \mathcal{P}_{Q^\perp}$, this implies that

$$\|\mathcal{P}_{Q^\perp} \mathcal{P}_\Omega \mathcal{P}_{Q^\perp}\| \leq 2\rho, \quad (83)$$

with high probability. It follows that

$$\|\mathcal{P}_{Q^\perp} \mathcal{P}_\Omega\| < 1/2, \quad (84)$$

with high probability, provided that ρ is sufficiently small and $\nu^2 p^3 \log m / n \leq C$, where C is a numerical constant.

References

- [1] J. Wright, A. Yang, A. Ganesh, Y. Ma, and S. Sastry, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, Feb 2009.
- [2] M. Fazel, H. Hindi, and S. Boyd, "Rank minimization and applications in system theory," in *American Control Conference*, June 2004.
- [3] C. Papadimitriou, P. Raghavan, H. Tamaki, and S. Vempala, "Latent semantic indexing: A probabilistic analysis," *Journal of Computer and System Sciences*, vol. 61, no. 2, Oct 2000.

- [4] C. Eckart and G. Young, “The approximation of one matrix by another of lower rank,” *Psychometrika*, vol. 1, pp. 211–218, 1936.
- [5] I. Jolliffe, *Principal Component Analysis*. Springer-Verlag, 1986.
- [6] E. Candès, X. Li, Y. Ma, and J. Wright, “Robust principal component analysis?” *Journal of the ACM*, vol. 58, no. 7, May 2011.
- [7] V. Chandrasekaran, S. Sanghavi, P. Parrilo, and A. Willsky, “Rank-sparsity incoherence for matrix decomposition,” *SIAM Journal on Optimization*, vol. 21, no. 2, pp. 572–596, 2011.
- [8] D. Hsu, S. M. Kakade, and T. Zhang, “Robust matrix decomposition with sparse corruptions,” *IEEE Transactions on Information Theory*, vol. 57, no. 11, pp. 7221–7234, 2011.
- [9] Z. Zhou, X. Li, J. Wright, E. Candès, and Y. Ma, “Dense error correction for low-rank matrices via principal component pursuit,” in *IEEE International Symposium on Information Theory*, 2010.
- [10] A. Ganesh, J. Wright, X. Li, E. Candès, and Y. Ma, “Dense error correction for low-rank matrices via principal component pursuit,” in *IEEE International Symposium on Information Theory*, 2010.
- [11] L. Wu, A. Ganesh, B. Shi, Y. Matsushita, Y. Wang, and Y. Ma, “Robust photometric stereo via low-rank matrix completion and recovery,” in *Asian Conference on Computer Vision*, 2010.
- [12] Z. Zhang, A. Ganesh, X. Liang, and Y. Ma, “TILT: Transform Invariant Low-rank Textures,” *International Journal of Computer Vision*, 2011. [Online]. Available: <http://dx.doi.org/10.1007/s11263-012-0515-x>
- [13] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma, “RASL: Robust Alignment by Sparse and Low-rank decomposition,” *To appear in IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011.
- [14] E. Candès and B. Recht, “Exact matrix completion via convex optimization,” *Found. of Comput. Math.*, 2008.
- [15] E. Candès and T. Tao, “The power of convex relaxation: Near-optimal matrix completion,” to appear in *IEEE Transactions on Information Theory*, 2009.
- [16] D. Gross, “Recovering low-rank matrices from few coefficients in any basis,” preprint, 2009.
- [17] X. Li, “Compressed sensing and matrix completion with constant proportion of corruptions,” 2011, available at <http://arxiv.org/abs/1104.1041>.
- [18] J. Wright, A. Ganesh, K. Min, and Y. Ma, “Compressive principal component pursuit,” 2012, submitted.
- [19] M. Ledoux, *The Concentration of Measure Phenomenon*. American Mathematical Society, 2001.
- [20] M. Rudelson and R. Vershynin, “Non-asymptotic theory of random matrices: extreme singular values,” in *Proc. of International Congress of Mathematicians*, 2010.
- [21] E. Candès and J. Romberg, “Sparsity and incoherence in compressive sampling,” *Inverse Problems*, vol. 23, no. 3, pp. 969–985, 2007.
- [22] J. Tropp, “User-friendly tail bounds for sums of random matrices,” *Foundations of Computational Mathematics*, 2011. [Online]. Available: <http://dx.doi.org/10.1007/s10208-011-9099-z>
- [23] R. Vershynin, “Introduction to the non-asymptotic analysis of random matrices,” 2011, available at <http://www-personal.umich.edu/~romanv/papers/non-asymptotic-rmt-plain.pdf>.